Benthic macroinvertebrates as bioindicators of water quality in a Kenyan highland river

Postgraduate Certificate in Ecological Survey Techniques

University of Oxford

Word count: 4,586

Table of contents

List of Figures

List of Tables

Abstract

Freshwater macrobenthos in the orders Ephemeroptera, Plecoptera and Trichoptera ('EPT' taxa) are known to be intolerant of poor water quality and are considered useful 'bioindicators' of environmental quality, being among the first macroinvertebrates to react to changes in an aquatic ecosystem. Many highland rivers in East Africa remain poorly studied, including the Mathioya, which is an important tributary of Kenya's longest river, the Tana. This survey sampled macrobenthos from 8 sites of low and high human disturbance along a short section of the Mathioya during the rainy season. Physico-chemical parameters used in water quality assessments were also measured, with associations between these and EPT taxa tested for statistical significance. Species abundance, richness and diversity were found to be generally higher at less disturbed sites, which also held a greater proportion of pollution intolerant EPT taxa. Further observations of the feeding ecology of the macrobenthos sampled appear to support assumptions described by the river continuum concept, with implications for river management guidelines in Kenya.

Acknowledgements

Many thanks to Dr Thomas Hesselburg, Dr Ada Grabowska-Zhang and Dr Cecilia Dahlsjö for their guidance whilst at Oxford, particularly for their support when my health deteriorated. Thanks to Laban Njoroge and the taxonomy team at National Museums of Kenya; to my fellow students; the staff of Kenya Fly Fishers' Camp for their hospitality whilst in the field, and above all to my wife Victoria and son Oliver for their patience and assistance with field research.

1. Introduction

Freshwater macroinvertebrate taxa in the orders Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies) ('EPT' taxa) are known to be intolerant of poor water quality (Lenat 1988; Lenat and Crawford 1994; Masese et al. 2009). Indeed, the 'EPT index', expressed as a percentage of these sensitive taxa to total taxa recorded, is among the most widely used indicators of environmental conditions in riverine ecosystems worldwide (Barbour et al. 1999; Masese and Raburu 2017). Being intolerant to water pollution, EPT taxa are considered useful 'bioindicators' of environmental quality (Lenat and Crawford 1994) and will be among the first macroinvertebrates to react to changes in an aquatic ecosystem, e.g. increased organic pollution and sedimentation downstream of human disturbance.

In Kenya, biomonitoring of aquatic ecosystems is generally wanting due to limited financial devotion, lack of technical capacity and limited guides on biomonitoring (Aura et al. 2021). Research over the last decade has focussed on filling this research gap to develop datasets, guidelines and bioindicator indices for East African taxa (e.g. Minaya et al. 2013; Mbaka et al. 2014; Masese and Raburu 2017). Indeed, many such systems remain poorly studied; the Mathioya River, which drains the eastern escarpment of the Aberdares mountain range in the central highlands of Kenya (Fig. 1), is one of them.

Fig. 1: Map of Kenya with study area indicated by red box: these central highlands are shown in greater detail in Fig. 2.

The Mathioya river (Fig. 2) originates in pristine, closed-canopy forest within the Aberdares national park, before entering a riparian gallery forest intersected by agricultural areas (primarily tea plantations) downstream. In its upper reaches, it is impounded by a dam, making for an interesting investigation with reference to the river continuum concept (Vannote et al. 1980), which describes spatial variability in the feeding ecology of stream macroinvertebrates vis-à-vis river morphology.

This study seeks to survey the macroinvertebrate community of the Mathioya at sites of lower and higher anthropogenic disturbance along a small stretch of its upper catchment.

An assumption to be tested is that agricultural activities close to the river cause sediment and pollutants to enter the watercourse, red ucing habitat quality for pollution intolerant taxa in the orders Ephemeroptera, Plecoptera and Trichoptera. More intact sites, of lower human disturbance, are conversely assumed to better maintain habitat quality, wherein one would correspondingly expect to record a higher EPT index.

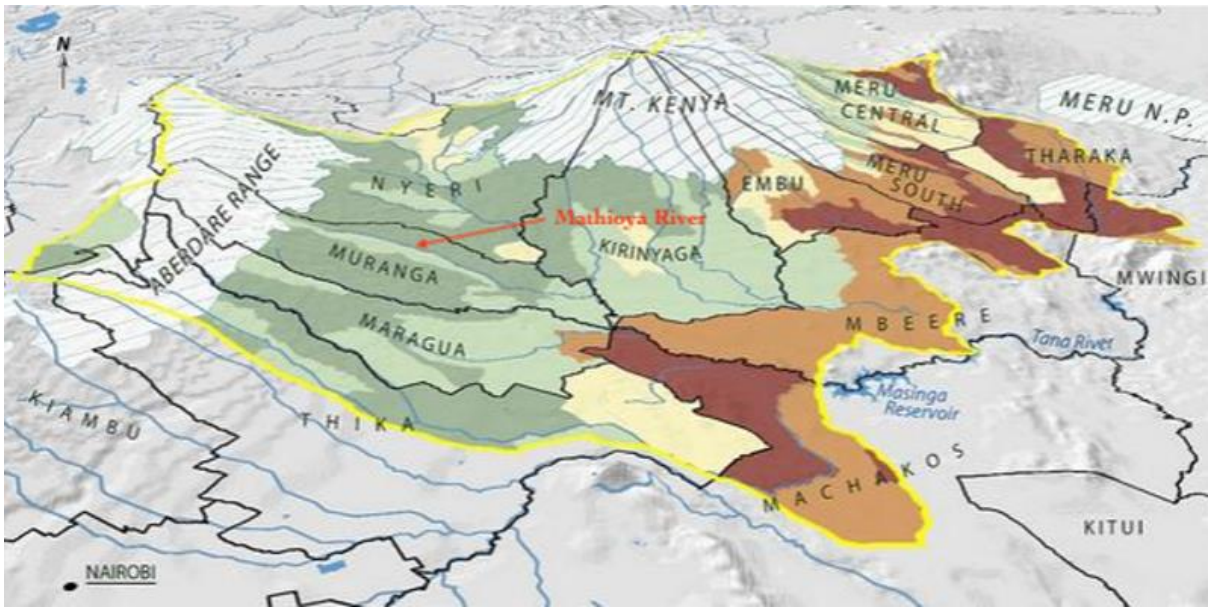


Fig. 2: 3D relief of the Tana River basin, showing the Mathioya (centre, indicated by red arrow) draining the eastern Aberdares catchment within the Kenyan highlands.

2. Literature review

Macrobenthos

Aquatic macroinvertebrates are invertebrates large enough (>0.5 mm) to be seen with the naked eye. Most are found within the bottom sediments of waterbodies and hence are often referred to as benthic macroinvertebrates. Examples include species of Mollusca and Annelida, which spend their entire lifetime in water, and aquatic insects of the orders Ephemeroptera, Plecoptera, Trichoptera and Diptera, whose immature stages (i.e. eggs, larvae and pupae) are aquatic, whereas subsequent adult stages are terrestrial (Jacobsen et al. 2008; Ochieng et al. 2019).

Adult stages of aquatic insects usually live for only a few days to a few weeks, whereas their larval or nymphal stages may last for a few months to over a year (Jacobsen et al. 2008). Benthic macroinvertebrate communities ('macrobenthos') are useful proxies for the health of aquatic ecosystems (Wenn 2008) owing to the sensitivity of their different life stages to environmental stressors (Hutchinson et al. 1998), as well as their relatively long lifespan, which enables them to integrate the effects of short and long-term variations into 'biorecords' of environmental change (Pratt and Coler 1976).

Anthropogenic impacts

Human activities, such as discharge of industrial effluents, agricultural intensification and urban runoff, have detrimentally affected the quality of rivers the world over (Laaffat et al. 2015). Contaminants released into waterbodies pose a threat to the supply of drinking water for humans and wildlife, whilst negatively affecting habitat conditions for a range of aquatic organisms, including fish (Santos et al. 2014). Reduced water quality thus impairs ecosystem health and services, compromises human needs, and hinders sustainable development (Lopez et al. 2013; Kostyla et al. 2015).

This is especially true for Kenya, where concern over water pollution in the nation's rivers has been raised since the 1990s (Githinji et al. 2019; Robert et al. 2020). Literature on the ecology, water quality and anthropogenic impacts of the Mathioya River is extremely limited, despite the important agricultural activities along its banks and it being a major tributary to the Tana, the longest river in Kenya, itself supporting a series of key reservoirs and major hydroelectric dams.

Research framework

Previous studies in Kenya have linked river pollution to surrounding anthropogenic activities (e.g. Masese and McClain 2012). Mbaka et al. (2014) linked livestock grazing and water abstraction for agriculture to a decline in habitat quality in the Honi and Naro Moru rivers, whilst Njuguna et al. (2017) recorded poor water quality along the length of the Nairobi river,

likewise attributing observed deterioration to industrial effluents, domestic sewage, agricultural activities and solid waste disposal.

Water quality may be assessed by researchers in different ways (Zahedi et al. 2017). Some assessments are based on a comparison of observed parameter values against established water quality guidelines for different uses or purposes (e.g. Minaya et al. 2013). This 'physico-chemical' approach is useful to identify sources of contamination has less utility when comparing sites with variable water impacts (Debels et al. 2005). A Water Quality Index (WQI), on the other hand, describes one of several methods that allow sites to be compared by reducing a set of variables into a single dimensionless number to rank water quality from poor to good, such that results are readily interpreted (Hosseini-Moghari et al. 2015; Robert et al. 2020).

This study uses a taxonomic, community based WQI (the EPT index) to assess water quality in the Mathioya river. Justification for this is threefold: (1) the method examines just three orders of insects that are easily sorted and identified, without the need for expert taxonomy; (2) the approach has been applied successfully across temperate zones for decades (e.g. Lenat 1988; Herbst et al. 2018), yet with a paucity of studies conducted in tropical systems (cf. Masese et al. 2009; Kripa et al. 2013), and (3) there have been recent calls for its increased application and refinement for sampling East African rivers (Mbaka et al. 2014; Masese and Raburu 2017), presenting the opportunity to contribute to the nascent literature thereon.

River Continuum Concept

According to the river continuum concept (Vannote et al., 1980), headwaters of river systems are typically small, shaded streams where allochthonous inputs (terrestrial material comprising leaves, sticks and other organic matter) of coarse particulate organic matter (CPOM) are a necessary resource for 'shredders' – one of several Functional Feeding Groups (FFGs) among the macrobenthos – having mouthparts adapted for feeding on this terrestrial plant litter.

Within lower reaches of a river system, where the river tends to be wider (Fig. 3), with reduced canopy cover from trees, the relative proportion of CPOM is lower, having been replaced by fine particulate organic matter (FPOM), which is produced largely from the feeding habits of the upstream shredders. This then becomes the food source for a variety of 'collectors' that have mouthpart adaptations for sweeping up FPOM from depositional areas or crevices where it has settled or been entrained (Cummins 2018). The proportion of a third group, predators, generally remains constant in both zones, these species being adapted to catch and consume live prey rather than feeding on material of terrestrial origin (Vannote et al., 1980).

The section of the Mathioya surveyed in the present study provides an opportunity to test the river continuum concept on a small scale, since the river here, whilst still in its upper reaches, has been impounded by a sizeable dam. This allows us to test the tenets of the concept by assessing the FFGs of macroinvertebrates sampled at different sites along the

river, with potential inferences drawn as to the relative contribution of FPOM from human activities at sites of high disturbance.

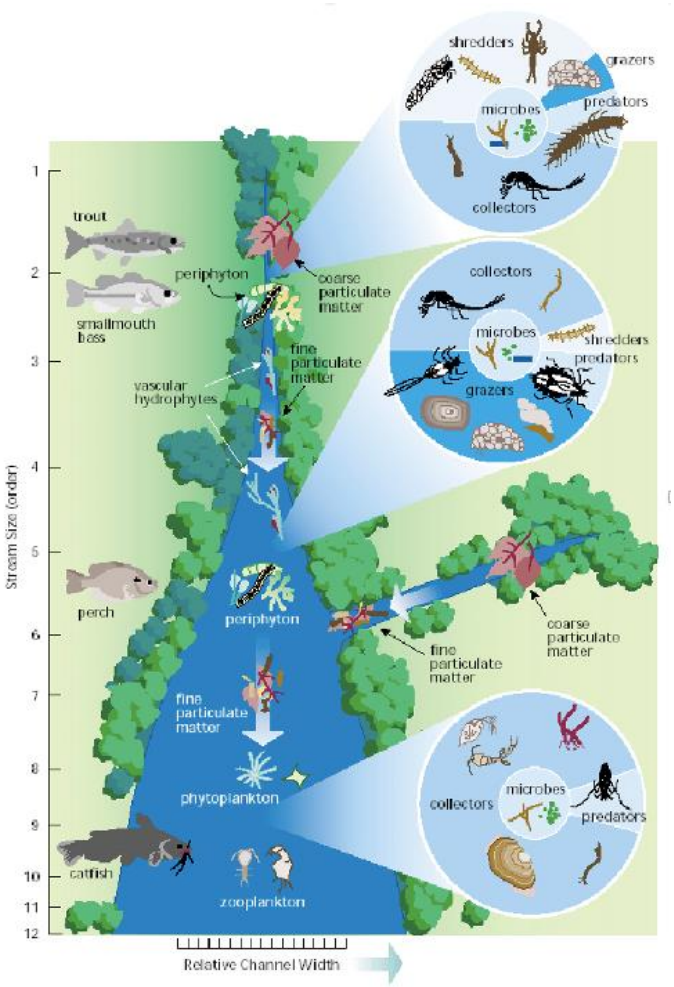


Fig. 3: Illustration of the river continuum concept and functional feeding groups among macrobenthos (from Vannote et al., 1980).

Indeed, research frameworks such as this, applying coarse-level taxonomy for macroinvertebrate assessments, have been shown to be sufficient for detecting environmental conditions even at small scales, particularly in environmental pollution and perturbation studies (Kaaya et al., 2015), and are increasingly incorporated into basic ecological investigations, including a recent investigation of highland African streams in Zimbabwe (Dalu et al. 2017).

3. Aims & research questions

The aim of the survey, then, is to ascertain whether differences can be observed between benthic macroinvertebrates sampled from sites of low vs. high disturbance along the Mathioya river. Sites of high disturbance are characterised by agricultural and other human activities close to the riverbank (predominantly tea cultivation: Fig. 4), whereas site of lower disturbance are those within forested sections of the river with intact, indigenous riparian flora.

Fig. 4: Steep slopes under tea cultivation abutting the banks of the Mathioya river.

The general assumption is that sites of lower disturbance will support a higher proportion of pollution-intolerant taxa (i.e. EPT) relative to more disturbed sites, in agreement with findings of van Biervliet et al. (2009) who conducted similar surveys among tea plantations in Tanzania, as well as other studies published for temperate zones (e.g. Herbst et al. 2018).

This assumption will be tested by addressing 3 research questions:

Q1: Is there evidence of variation in benthic communities between sites of low and high disturbance?

The null hypothesis (H_0) is that there is no variation in the macrobenthos between sites of low and high disturbance; the alternative hypothesis (H_1) is that there is variation between the sites.

Q2: Is there a significant association between EPT taxa and physico-chemical parameters of water quality?

H_0 = there is no significant association between EPT taxa and physico-chemical parameters; H_1 = there are differences between the two sets of variables.

Q3: Is there evidence of variation in Functional Feeding Group (FFG) between sites of low and high disturbance?

H_0 = there is no evidence of variation in FFG between sites of low and high disturbance; H_1 = there is evidence of variation between the sites.

By investigating a poorly studied river system, an underlying objective of the survey is to contribute novel data towards establishing refined bioindicator guidelines for tropical water quality monitoring in Kenya, a need highlighted by several research teams currently active in the field (Masese et al. 2014; Masese and Raburu 2017; Ochieng et al. 2019).

4. Methodology

Site selection

During a scoping visit to the Northern Mathioya in the weeks preceding data collection, a 3.5 km stretch was surveyed to identify riffle zones within locations of low (L) and high (H) human disturbance along the river. A total of 12 locations, where access to the river was considered safe, were encountered and their coordinates logged using a GPS unit. From these data points, a total of 8 (4 x L; 4 x H) were randomly selected (using Excel) for data collection (Fig. 5).

Fig. 5: Sites of low (L) and high (H) disturbance sampled along the Northern Mathioya; river runs downstream from left to right; note the dam in the centre of the image above site H3; star denotes base camp.

Sampling was then conducted during the long rainy season (March-April), when impacts from human disturbance (e.g. sedimentation from agricultural runoff) were likely to be amplified. This decision was made in order to increase the chances of detecting differences in the macrobenthos over a fairly short section of river, given the time and resource-limited nature of project[1].

Sites of low disturbance were characterised as having an intact riparian flora, often within a forested section of the river (in some cases with canopy covering the channel), generally away from human activities such as farming (Fig. 6). On the other hand, sites of higher disturbance had less intact riparian zones, reduced canopy cover and banks that had either been cleared for grazing or cultivation of cash crops (primarily tea), these areas being associated with agriculture.

Fig. 6: A site of low anthropogenic disturbance along the Mathioya: note the tree-lined banks and intact riparian vegetation.

Headwater sections of rivers such as these host the majority of their macroinvertebrate fauna within shallow riffle zones: short sections where shallow, fast-flowing water is agitated by rocks (Fig. 7). Riffles zones were chosen as they are known to be productive sites of EPT and other macroinvertebrate taxa (Herbst et al. 2018), are readily identifiable by amateur surveyors (Storey et al. 1991) and are generally safely accessible for wading (being typically <1 m deep).

[1] Given the opportunity to do so, the author would compare the results of this study with those from the same surveys repeated during the dry season, in order to examine the effects of any seasonal variation on macroinvertebrate assemblages.

Fig. 7: Typical riffle zone along the Mathioya, ideal for sampling macroinvertebrates from the riverbed.

Kick-sampling

During subsequent data collection (March-April 2022), these 8 sites were revisited, starting downstream at site L4 and working upstream towards L1 to reduce impacts from the surveyor's disturbance. At each site, a 30 m section of riffle habitat (starting from the head of the 'run', which follows a riffle) was marked out with tape for kick-sampling.

Kick-sampling is a useful method for collecting benthic macroinvertebrates from shallow waters such as these riffle zones. The technique involves agitating the stones or sediment of a river by foot and collecting the sample in a sturdy hand net that is held downstream (Fig. 8). Samples were collected using a D-frame net (1 mm mesh size; 250 mm wide, 0.30 m deep), with the surveyor moving in a zig-zag fashion upstream along the 30 m section for a standardized sampling period of 3 minutes, following Storey et al. (1991). Stones and small sticks were washed off carefully into the net before the sample was rinsed out into a tray full of river water on the bank for the purposes of sorting and identification.

During the scoping visit, the opportunity was taken to pilot the kick-sampling method. It was determined that a total of 5 points within a 30 m section could comfortably be sampled within the 3-minute window, equivalent to 1 m^2 of riverbed.

Fig. 8: Kick-sampling a site (left) and taking measurements of electrical conductivity (right).

Physico-chemical parameters

After kick-sampling in this way at each site (i.e. post-disturbance), the average width and depth of the riffle was recorded at 10 m intervals along the section to determine the mean cross-sectional area (*CSA*):

$$CSA\ (m^2) = Width\ (m) \times Depth\ (m)$$

River velocity (*v*) was then determined by recording the average time taken for an orange[2] to travel down the 30 m section over 3 repeats, a simple and cost-effective method for the amateur surveyor.

Discharge (*Q*) was then calculated as:

$$Q\ (m^3/s) = v\ (m/s) \times CSA\ (m^2)$$

[2] An orange is considered a suitable object to use because it has enough buoyancy to float just below the water surface where maximum velocity typically occurs (Storey et al. 1991).

Additional physico-chemical parameters relating to water quality – electrical conductivity (EC), total dissolved solids (TDS), temperature (°C) and pH – were recorded using a simple handheld device (Semlos ECM: Fig. 8) at a fixed depth of 10 cm below the water surface in the centre of the riffle, following Storey et al. (1991).

Taxonomy

Back at base camp, pooled samples were first hand sorted into their respective orders. Where necessary, specimens were fixed in 10% formaldehyde and then identified to the level of family *in situ* by the author using an Ceti Max II Trino microscope (Fig. 9) with reference to Gerber and Gabriel (2002). Challenging specimens were set aside and later sent to taxonomists at the National Museums of Kenya who assisted in further identification.

Fig. 9: Morphological identifiers supporting classification into families (left panel), setting up the microscope, from which these images were taken (right panel).

Species richness was determined from the number of different families identified at each site. Species abundance (per m^2) was determined by multiplying the sum of different taxa collected at each site by the area sampled. Species diversity (*H*) was calculated using the Shannon-Wiener index:

$$H = -\Sigma(Pi) \times ln(Pi)$$

where *Pi* = the proportion of individuals of each species.

The EPT index was then calculated to show the proportion of taxa from the orders Ephemeroptera, Plecoptera and Trichoptera relative to all taxa recorded, after Barbour et al. (1999).

Finally, the Functional Feeding Group (FFG) of each family was assigned after further desk study with reference to Cummins (2018) and other published literature for Kenya (Masese et al. 2014), allowing for differentiation of specimens into 4 FFGs: shredders, collectors, scrapers and predators.

Statistical analyses

Tests for significance between the variables were performed in R in order to address the 3 research questions set out above. Data were checked for normality before applying the student's t-test (e.g. for variance between sites of L vs. H disturbance), multivariate analysis (e.g. of physico-chemical parameters) or ANOVA (e.g. to tests for variation among FFGs) set against 95% confidence intervals.

5. Results

Site characteristics

The 8 sites surveyed covered a length of 3,125 m and ranged in altitude from 2,017 m (site L1) down to 1,922 m (L4), a height difference of 95 m. The mean river width and depth at these riffle habitats was 12.7 (±0.6) m and 0.6 (±0.03) m respectively. Mean discharge was determined to be 11.6 (±0.9) m^3/s (Table 1).

Mean results for physico-chemical parameters across the survey sites were as follows: water temperature 19.2(±0.2) °C, pH 8.4(±0.1), electrical conductivity (EC) 134.6(±39.2) µs/s and total dissolved solids (TDS) 64.9(±19.4) mg/l. Whilst temperature and pH values were similar between sites of low (19.4±0.3 °C and 8.3±0.1 respectively) and high (19.1±0.2 °C and 8.5±0.2 respectively) disturbance, EC and TDS were found to increase at site of high disturbance (169.2±67.1 µs/s and 82.5±34.5 mg/l) relative to sites of low disturbance (100.1±43.2 µs/s and 47.4±19.1 mg/l).

Parameter	L1	H1	L2	H2	H3	L3	H4	L4	Means
Coordinates	0°36'50"S, 36°51'22"E	0°36'57"S, 36°51'30"E	0°37'06"S, 36°41'51"E	0°37'12"S, 36°51'44"E	0°37'11"S, 36°51'57"E	0°37'13"S, 36°52'22"E	0°37'19"S, 36°52'29"E	0°37'22"S, 36°52'40"E	
Altitude (m)	2,017	1,994	1,984	1,977	1,962	1,940	1,935	1,922	1,966.38 (±11.51)
Width (m)	10.2	11.6	11.2	15.2	14.9	12.4	13.1	12.9	12.69 (±0.61)
Depth (m)	0.56	0.53	0.64	0.49	0.6	0.65	0.71	0.58	0.60 (±0.03)
CSA (m^2)	5.71	6.15	7.17	7.45	8.94	8.06	9.3	7.48	7.53 (±0.44)
v (m/s)	1.52	1.64	1.41	1.21	1.42	1.45	1.75	1.86	1.53 (±0.07)
Q (m^3/s)	8.68	10.08	10.11	9.01	12.69	11.69	16.28	13.92	11.56 (±0.93)
Temp (°C)	18.9	19.3	19	20.4	19.6	18.6	19.2	18.9	19.24 (±0.10)
pH	8.2	8.4	8.1	8.4	8.9	8.1	8.6	8.4	8.39 (±0.20)
EC (µs/s)	40.2	84.8	48.7	226.5	361.9	85.4	159.3	70.1	134.61 (±39.20)
TDS (mg/l)	20.5	40.8	24.9	103.2	184.7	45.3	63.2	36.9	64.94 (±19.43)
	Downstream →								

Table 1: Physico-chemical parameters measured at each site of low (L) and high (H) disturbance, presented in downstream order of descending altitude, standard errors of mean values shown in parentheses.

Macrobenthos

A total of 16 families within 9 orders were identified from the specimens collected across the 8 sites (Table 2). The family Baetidae (order Emphemeroptera: Fig. 10) had the highest mean abundance across all sites (2015.6±130.9 individuals/m^2: Fig. 11) whereas Potamonautidae (Decapoda: Fig. 12) had the lowest (8.3±4.4 ind./m^2). Species abundance was highest at L2 (6,561.7 taxa/m^2) and lowest at L4 (3,188.2 taxa/m^2). Total abundance was slightly higher at sites of low disturbance (5,100.8±807.4 taxa/m^2) than sites of higher disturbance (4,578.7±316.5 taxa/m^2).

Fig. 10: Mayfly larvae of the family Baetidae, frequently encountered during the survey.

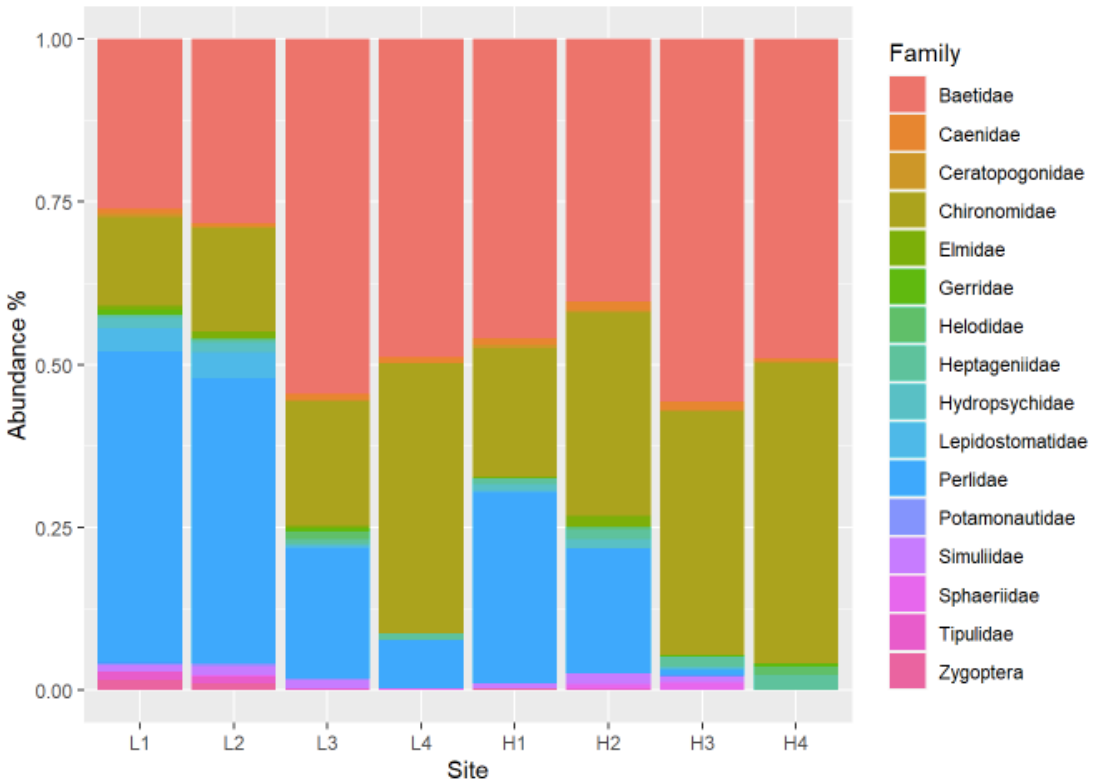


Fig. 11: Abundance of the 16 families sampled across the 8 sites.

The site with the greatest species richness (15) was L2, with L4 having the lowest (6). Both species richness and diversity (*H*) were greater at sites of lower (12±2 and 1.32±0.1 respectively) than higher (10±1 and 1.16±0.1) disturbance. Diversity was greatest at L1 (*H* = 1.51), at the highest altitude, and lowest at H4 (*H* = 0.91), the site of lowest altitude; these two sites also had the highest and lowest EPT indices, with scores of 80.3% and 51.9% respectively.

Sites above the dam had greater species abundance, richness and diversity (5,722.8±438.6, 13.0±0.9 and 1.3±0.1 respectively) than sites below the dam (4,309.9±292.0, 9.6±1.5 and 1.2±0.1 respectively).

Fig. 12: Freshwater crab of the family Potamonautidae, probably *Potamonautes sp.*, a shredder.

Order	**Family**	**L1**	**L2**	**L3**	**L4**	**H1**	**H2**	**H3**	**H4**
Ephemeroptera	Baetidae	1645.2	1853.5	2366.3	1556.6	2462.1	1876.6	2488.8	1875.3
	Caenidae	45.8	33.2	43.8	31.2	54.6	63.7	63.2	23.4
	Heptageniidae	23.1	25.6	34.1	33.7	44.6	74.2	67.8	85.4
Plecoptera	Perlidae	3014.4	2866.5	865.4	234.4	1567.5	890.4	43.4	0
Trichoptera	Hydropsychidae	108.7	98.4	12.3	0	57.4	66.4	19.2	0
	Lepidostomatidae	225.4	265.9	14.7	0	16.2	0	0	0
Diptera	Ceratopogonidae	34.4	18.1	0	0	28.7	15.6	0	0
	Chironomidae	856.6	1043.4	835.2	1321.1	1056.2	1455.5	1677.2	1765.4
	Simuliidae	64.2	89.7	55.6	0	44.6	79.2	45.4	0
	Tipulidae	78.3	67.5	13.3	0	0	12.5	0	0
Odonata	Zygoptera	102.8	68.4	0	0	14.5	0	0	0
Coleoptera	Elmidae	44.7	62.1	19.7	0	15.5	75.3	0	0
	Helodidae	0	19.8	52.1	0	0	16.8	0	48.1
Hemiptera	Gerridae	43.6	0	19.2	0	0	0	11.2	21.4
Decapoda	Potamonautidae	21.3	31.8	13.2	0	0	0	0	0
Sphaeriida	Sphaeriidae	0	17.8	0	11.2	0	32.7	54.4	4.2
	Total abundance	6,308.5	6,561.7	4,344.9	3,188.2	5,361.9	4,658.9	4,470.6	3,823.2
	Species richness	14	15	13	6	11	12	9	7
	EPT Total	5,062.6	5,143.1	3,336.6	1,855.9	4,202.4	2,971.3	2,682.4	1,984.1
	EPT %	80.25	78.38	76.79	58.21	78.38	63.78	60.00	51.90
	Species diversity (*H*)	1.51	1.49	1.26	1	1.29	1.46	0.99	0.91

Table 2: Summary data of taxa collected at the 8 sites showing abundance, species richness, species diversity and percentage of EPT taxa recorded.

In terms of Functional Feeding Groups, the most abundant guild across all sites surveyed were Predators, with a mean representation of 65.6(±3.4)% of the survey pool, followed by Collectors 31.6(±3.5)%, Shredders 2.3(±0.6)% and Scrapers 0.6(±0.2)%. Predators were dominated by chironomid larvae (Diptera: Chironomidae); Collectors by mayfly larvae in the family Baetidae (Ephemeroptera) (Fig. 11); Shredders by caddisfly larvae in the family Lepidostomatidae (Trichoptera) and Scrapers by mayfly larvae in the family Heptageniidae (Ephemeroptera).

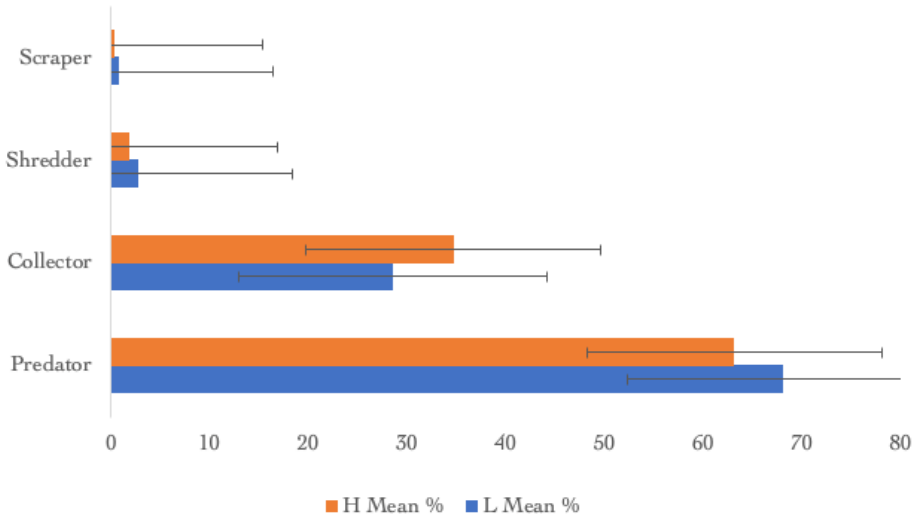


Fig. 12: Mean percentages of FFGs sampled at sites of high (H) and low (L) disturbance showing standard error bars.

Mean proportional representation of FFGs was found to be similar between sites of low and high disturbance (Fig. 12). However, sites below the dam had over twice as many (61.6%) collectors than those above the dam (26.5%) and, conversely, less than half the shredders (1.3%) relative to upstream sites (3.3%). The proportion of scrapers was <1% on either side of the dam, whilst predators were found in greater abundance above (69.5%) than below the dam (36.8%).

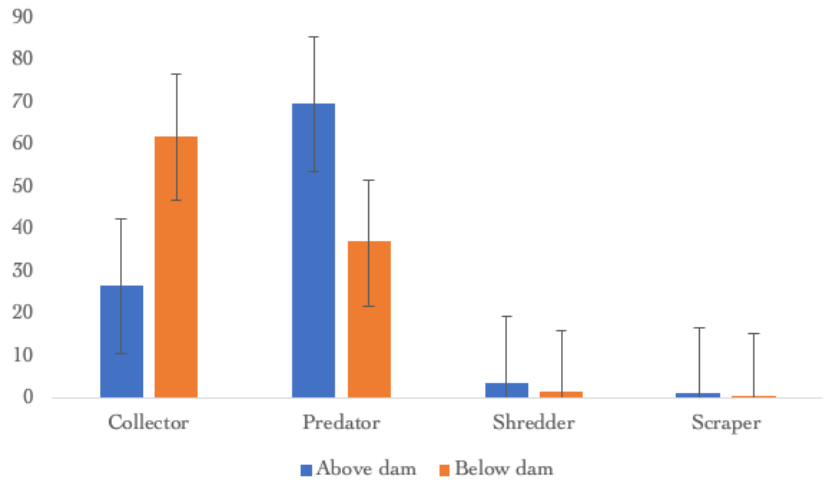


Fig. 13: Proportional representation of FFGs above and below the dam showing standard error bars.

Statistical analyses

Applying the student's t-test to the normally distributed data (supported by the Shapiro-Wilk test) comparing macrobenthos communities between sites of low and high disturbance revealed no significant differences (with a 95% confidence interval) for abundance ($t(3.9) = -0.602$, $p = 0.5$), species richness ($t(4.6) = -0.969$, $p = 0.4$), diversity ($t(6.0) = -0.869$, $p = 0.4$) or EPT index ($t(6.0) = -1.313$, $p = 0.2$).

Performing multivariate analyses on EPT indices and physico-chemical characteristics revealed no significant differences between sites (see Appendix), although correlation between EPT% and altitude was almost significant ($F(1,6) = 5.778$, $p = 0.053$).

Testing for variation among FFGs at each site was found to be highly significant ($F(7,24) = 63.42$, $p < 0.001$), although the difference between FFGs and sites of low and high disturbance was only weakly significant ($p = 0.09$).

6. Discussion

Macroinvertebrate assemblage

There was no significant ($p < 0.05$) variation in the macrobenthos sampled between sites of low and high disturbance (Q1) along the section of the Mathioya surveyed during this study, we can therefore accept the null hypothesis. That said, sites of lower disturbance were generally found to have greater abundance, species richness and higher diversity of taxa than more disturbed sites (whilst not statistically significant), in broad agreement with the findings of van Biervliet et al. (2009) who conducted a similar study among tea plantations in Tanzania. Likewise, Masese et al. (2014) recorded higher species richness in close-canopy forested streams, analogous to the sites of low disturbance sampled here. The reason for the lack of significant variation in the present study could be the small sample area (i.e. a short section of one river) relative to the 6 different rivers surveyed by van Biervliet et al. (2009).

Physico-chemical associations

No significant association was found between EPT indices and physico-chemical parameters of water quality (Q2), meaning we should again accept the null hypothesis. The correlation between EPT index and altitude was, however, almost significant ($p = 0.053$), in general agreement with the findings of Kasangaki et al. (2008), who found a similar correlation in high-altitude streams of Uganda, as well as Che Salmah et al. (2014) sampling macrobenthos in Malaysia, alongside others working in temperate biomes (e.g. Herbst et al. 2018). Perhaps if a greater altitudinal range had been sampled in this study, the association with EPT taxa would have become significant. Furthermore, the sampled area here did not include pristine forests of the upper reaches of the river (owing to access restrictions), wherein one would anticipate better habitat quality to support these pollution intolerant groups.

The similarity in temperature and pH between sites can perhaps likewise be explained by the minimal altitudinal variation between samples, as well as the short distance between them. However, the variance in EC and TDS measurements (common metrics across Water Quality Indices), which were on average twice as high at more disturbed sites, where lower species richness, diversity and fewer EPT taxa were recorded, lends general (if not statistically significant) support to arguments that runoff from agriculture in these headwater regions impairs both water and habitat quality as captured in recent literature (e.g. Githinji et al. 2019; Robert et al. 2020). In this case, the higher EC and TDS measurements might be presumed to originate from tea cultivation and livestock rearing on cleared riverbanks adjacent to the sample sites (Fig. 14), particularly during the wet season when these surveys were undertaken.

Fig. 14: Tea cultivation on steep slopes at a site of high disturbance

Functional Feeding Groups

Whilst the variation between FFGs and sites of low and high disturbance (Q3) was only weakly significant ($p = 0.09$), meaning we must once more accept the null hypothesis, it is noteworthy that variation among FFGs at each site was found to be highly significant ($p < 0.001$). The low species richness of the shredder guild (just three families) is perhaps surprising given the findings of Masese et al. (2014) who recently recorded a total of 19 in similar (i.e. 1,900-2,300 m) Kenyan highland streams – albeit the number of streams and study sites (20) was greater than in the present survey.

The significant difference between FFGs observed here is interesting in that sites below the dam had, on average, over twice as many collectors as those above the dam, but only half the abundance of shredders. Returning to the river continuum concept (RCC) of Vannote et al. (1980), these observations suggest that accumulation of FPOM from upstream shredder activity, as well as natural sedimentation within the dam, favours the dominance of collectors with mouthparts adapted to feed on this material. Observations made here, despite being over a short distance (3.5 km), thus appear to lend support to the RCC, albeit the widening of the river channel in this case was man-made (i.e. the result of impoundment) rather than geomorphological in nature.

Conversely, however, and contrary to the assumption that predators remain proportionally constant along the river system (Cummins 2018), predators were found in far greater abundance above rather than below the dam (Fig. 13). In this instance, the unusual result might be explained by the presence of trout in the dam, the Mathioya being a stocked river fishery, which are likely to prey preferentially upon the larger-bodied predator guild.

Study limitations

This survey was highly time and resource limited. Had this not been the case, the surveyor would have sampled further sites across a wider area and repeated these within the dry season to test for temporal as well as spatial variation. Access to pristine, upper reaches of

the river was not possible (being within a protected national park), nor was sampling the distant lower reaches deemed safe or practicable.

7. Conclusion

The macrobenthos comprises a great diversity of species, with a wide range of trophic levels and pollution tolerances (Kripa et al. 2013), limited movement and quick response to pollutants, including excess nutrients, sediment and other environmental stressors (Muralidharan et al. 2010) even over short distances, as can be observed in part from the present study.

Even small-scale, short-term surveys (*sensu* Kaaya et al., 2015) such as this can be seen to reveal an association between sensitive macrobenthos (EPT taxa) and anthropogenic disturbance (here, tea cultivation along steep slopes), lending support (however limited statistically) to arguments to protect the riparian zone from degradation. Indeed, environmental law in Kenya stipulates that a 30 m-wide reserve from the banks of its rivers must be remain free of clearance and cultivation in order to better protect the country's water resources (Aura et al. 2011). The data collected here adds, in a small way, to the scientific knowledge on which such legislation draws.

Developing countries, including Kenya, are grappling with increasing water quality deterioration as urban and agricultural populations expand and untreated effluent and runoff find their way into lakes and rivers (Besada and Werner 2015). It is therefore vital that water quality is monitored through the regular assessment of selected chemical, biological and physical properties (Jang et al. 2013; Karkra et al. 2017) and that this information is used to inform better aquatic resources management (Ochieng et al. 2019).

The less-studied Mathioya, being a tributary of the Tana, the longest river in Kenya and an important hydrological resource supporting human development, is a case in point.

References

Aura, C.M., Raburu, P.O., Herrmann, J. 2011. Macro-invertebrates' community structure in Rivers Kipkaren and Sosiani, River Nzoia basin, Kenya. *Journal of Ecology and the Natural Environment* 3: 39-46.

Barbour, M.T., Gerritsen, J., Snyder, B.D., Stribling, J.B. 1999. *Rapid bioassessment protocols for use in wadeable streams and rivers*. Periphyton, benthic macroinvertebrates and fish. 2nd edition. US Environmental Protection Agency, Office of Water, Washington, DC.

Besada, H., Werner, K. 2015. An assessment of the effects of Africa's water crisis on food security and management. *International Journal of Water Resources Development* 31: 120-133.

Che Salmah, M.R., Al-Shami, S.A., Abu Hassan, A., Madrus, M.R., Huda, A.N. 2014. Distribution of detritivores in tropical forest streams of peninsular Malaysia: role of temperature, canopy cover and altitude variability. *International Journal of Biometeorology* 58: 679-690.

Cummins, K.W. 2018. *Functional Analysis of Stream Macroinvertebrates*. In: Gökçe, D. (Ed.), Limnology - Some New Aspects of Inland Water Ecology. IntechOpen, London.

Dalu, T., Wasserman, R.J., Tonkin, J.D., Alexander, M.E., Dalu, M.T.B., Motitsoe, S.N., Manungo, K.I., OBepe, O., Dube, T. 2017. Assessing drivers of benthic macroinvertebrate community structure in African highland streams: An exploration using multivariate analysis. *Science of The Total Environment* 601-602: 1340-1348.

Gerber A., Gabriel, M.J.M. 2002. *Aquatic Invertebrates of South African Rivers - Illustrations version 2*. Pretoria, South Africa: Department of Water Affairs and Forestry, Institute for Water Quality Studies.

Githinji, M.W., Mwaura, F., Wamalwa, J. 2019. Land use and water pollution along the altitudinal gradient of the Likii River, Laikipia County, Kenya. *Journal of Environment Pollution and Human* Health 7: 39-52.

Herbst, D., Cooper, S., Medhurst, R., Wiseman, S., Hunsaker, C. 2018. A comparison of the taxonomic and trait structure of macroinvertebrate communities between the riffles and pools of montane headwater streams. *Hydrobiologia* 820: 115-133.

Hutchinson, T.H., Solbe, J., Kloepper-Sams, P.J. 1998. Analysis of the Acetoc Aquatic Toxicity (EAT) Database iii – Comparative Toxicity of Chemical Substances to different Life Stages of Aquatic Organisms. *Chemosphere* 36: 129–142.

Jacobsen, D., Cressa, C., Mathooko, J.M., Dudgeon, D. 2008. Macroinvertebrates: Composition, Life Histories and Production. In: Dudgeon, D. (Ed.), Tropical Stream Ecology. Academic Press, London.

Jang, C.S., Chen, S.K., Kuo, Y.M. 2013. Applying indicator-based geostatistical approaches to determine potential zones of groundwater recharge based on borehole data. *Catena* 101: 178-187.

Kaaya, L.T., Day, J.A., Dallas, H.F. 2015. Tanzania River Scoring System (TARISS): a macroinvertebrate-based biotic index for rapid bioassessment of rivers. *African Journal of Aquatic Science* 40: 109-117.

Karkra, R., Kumar, P., Bansod, B.K., Bagchi, S., Sharma, P., Krishna, C.R. 2017. Classification of heavy metal ions present in multi-frequency multi-electrode potable water data using evolutionary algorithm. *Applied Water Science* 7: 3679-3689.

Kasangaki, A., Chapman, L.J., Balirwa, J. 2008. Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. *Freshwater Biology* 53:681-697.

Kostyla, C., Bain, R., Cronk, R., Bartram, J. 2015. Seasonal variation of faecal contamination in drinking water sources in developing countries: a systematic review. *The Science of the Total Environment* 514: 333-343.

Kripa, P.K., Prasanth, K.M., Sreejesh, K.K., Thomas, T.P. 2013. Aquatic Macroinvertebrates as Bioindicators of Stream Water Quality: A Case Study in Koratty, Kerala, India. *Research Journal of Recent Sciences* 2:217-222.

Laaffat, J., Ouazzani, N., Mandi, L. 2015. The evaluation of potential purification of a horizontal subsurface flow constructed wetland treating greywater in semi-arid environment. *Process Safety and Environmental Protection* 95: 86-92.

Lenat, D.R. 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. *Journal of the North American Benthological Society* 7: 222–233.

Lenat, D.R., Crawford, J.K. 1994. Effects of land use on water quality and aquatic biota of three North American piedmont streams. *Hydrobiologia* 294: 185–199.

Lopez, A.B., Ramos, L.J.A., Santa, C.G., Moran Ramirez, J., Carranco Lozada, S.E., Noyola Medrano, M.C., Pineda Martinez, L.F. 2013. Water poverty index assessment in semi-arid regions: the case of San Luis Potosi Valley (Mexico). *Revista Internacional de Contaminación* Ambienta 29: 249-260.

Masese, F.O., McClain, M.E. 2012. Trophic resources and emergent food web attributes in rivers of the Lake Victoria Basin: a review with reference to anthropogenic influences. *Ecohydrology* 5: 685-707.

Masese, F.O., Muchiri, M., Raburu, P.O. 2009. Macroinvertebrate assemblages as biological indicators of water quality in the Moiben River, Kenya. *African Journal of Aquatic Science* 34: 15-26.

Masese, F., Kitaka, N., Kipkemboi, J., Gettel, G., Irvine, K., McClain, M. 2014. Macroinvertebrate functional feeding groups in Kenyan highland streams: Evidence for a diverse shredder guild. *Freshwater science* 33: 435-450.

Masese, F.O., Raburu, P.O. 2017. Improving the performance of the EPT Index to accommodate multiple stressors in Afrotropical streams. *African Journal of Aquatic Science* 42: 219-233.

Mbaka, J.G., M'Erimba, C.M., Thiongo, H.K., Mathooko, J.M. 2014. Water and habitat quality assessment in the Honi and Naro Moru Rivers, Kenya, using benthic macroinvertebrate assemblages and qualitative habitat scores. *African Journal of Aquatic Science* 39: 361–368.

Minaya, V., McClain, M.E., Moog, O., Omengo, F., Singer, G.A. 2013. Scale-dependent effects of rural activities on benthic macroinvertebrates and physico-chemical characteristics in headwater streams of the Mara River, Kenya. *Ecological Indicators* 32: 116-122.

Muralidharan, M., Selvakumar, C., Sundar, S., Raja, M. 2010. Macroinvertebrates as Potential Indicators of Environmental Quality. *Indian Journal of Biotechnology* 1: 23-28.

Njuguna, S.M., Yan, X., Gituru, R.W., Wang, Q., Wang, J. 2017. Assessment of macrophyte, heavy metal, and nutrient concentrations in the water of the Nairobi River, Kenya. *Environmental Monitoring and Assessment* 189: 454.

Ochieng, H., Okot-Okumu, J., Odong, R. 2019. Taxonomic challenges associated with identification guides of benthic macroinvertebrates for biomonitoring freshwater bodies in East Africa: A review. *African Journal of Aquatic Science* 44: 113-126.

Pratt, J.M., Coler, R.A. 1976. A procedure for the routine biological evaluation of urban runoff in small rivers. *Water Research* 10: 1019-1025.

Robert, G.K., Onyari, C.N., Mbaka, J.G. 2020. Development of a Water Quality Assessment Index for the Chania River, Kenya. *African Journal of Aquatic* Science 46: 142-152.

Santos, J.C., Andrade, E.M., Araújo Neto, J.R., Meireles, A.C., Palácio, H.A. 2014. Land use and trophic state dynamics in a tropical semi-arid reservoir. *Revista Ciência Agronômica* 45: 35-44.

Storey, A.W., Edward, D.H., Gazey, P. 1991. Recovery of aquatic macroinvertebrate assemblages downstream of the Canning Dam, Western Australia. *Regulated Rivers: Research & Management* 6: 213-224.

van Biervliet, O., Wiśniewski, K., Daniels, J., Vonesh, J.R. 2009. Effects of Tea Plantations on Stream Invertebrates in a Global Biodiversity Hotspot in Africa. *Biotropica* 41: 469-475.

Vannote, R.L., Wayne Minshall, G., Cummins, K.W., Sedell, J.R., Cushing, C.E. 1980. The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130-137.

Wenn, C.L. 2008. Do freshwater macro-invertebrates reflect water quality improvements following the removal of point source pollution from Spen Beck, West Yorkshire? *Earth and Environment* 3: 369-406.

Appendices

Statistical analyses

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

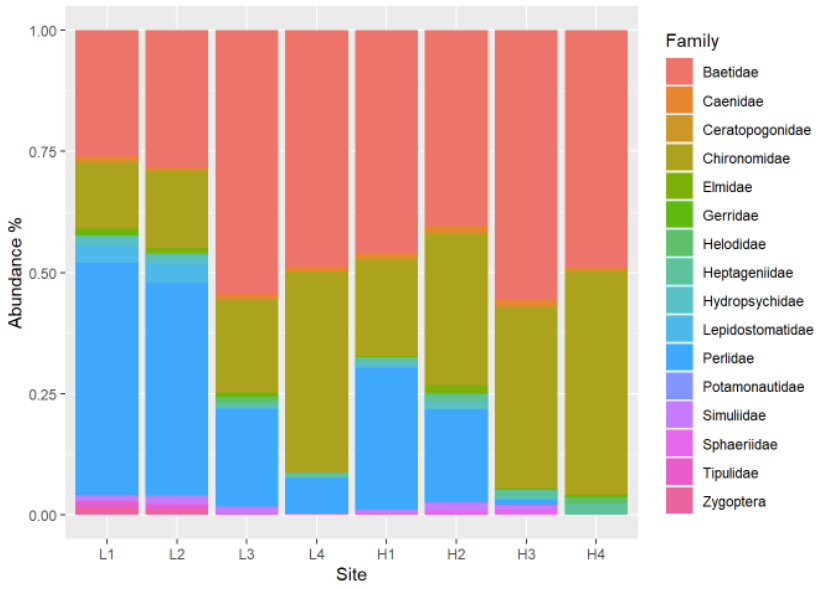
```
## v tibble  3.1.7.9000     v dplyr   1.0.9
## v tidyr   1.2.0          v stringr 1.4.0
## v readr   2.1.2          v forcats 0.5.1
## v purrr   0.3.4
```

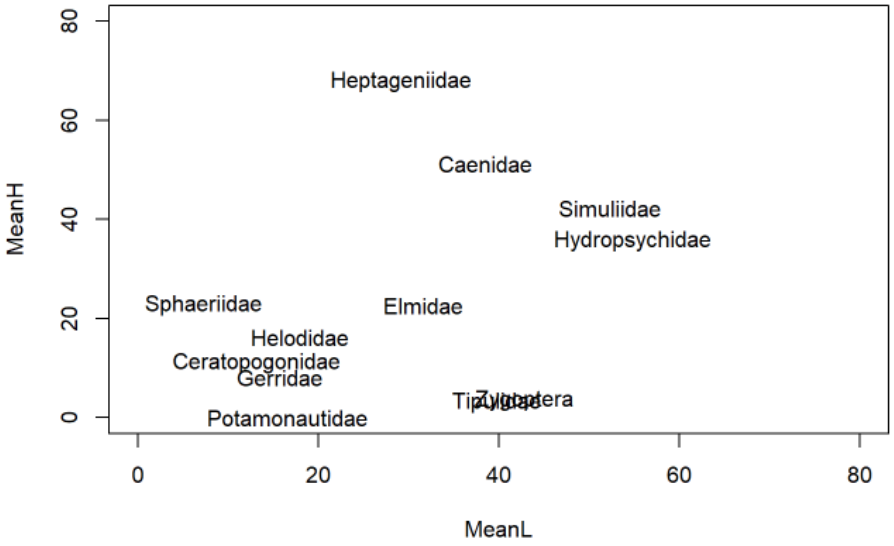
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
##
## Attaching package: 'rstatix'
```

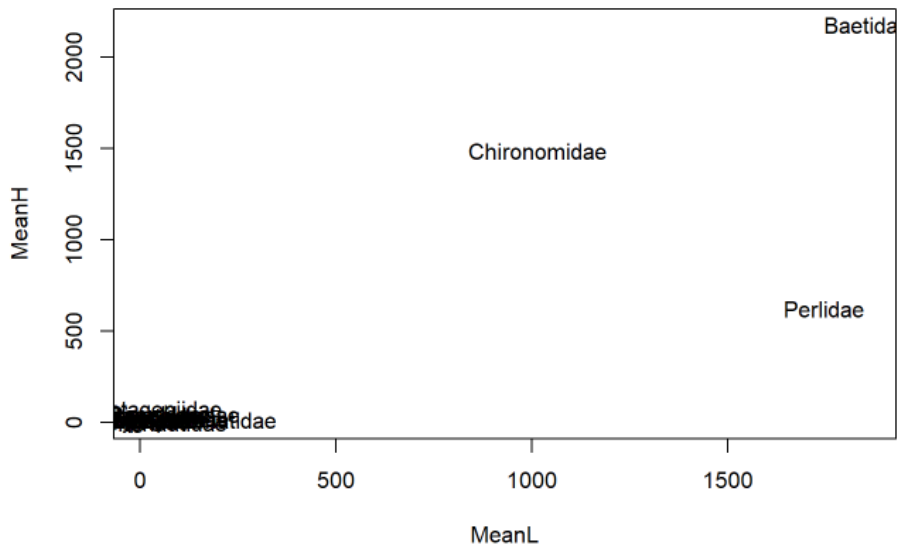
```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## Using Family, Group as id variables
```


1.00
0.75
0.50
0.25
0.00
Abundance %
L1
L2
L3
L4
H1
H2
H3
H4
Site
Family
Baetidae
Caenidae
Ceratopogonidae
Chironomidae
Elmidae
Gerridae
Helodidae
Heptageniidae
Hydropsychidae
Lepidostomatidae
Perlidae
Potamonautidae
Simuliidae
Sphaeriidae
Tipulidae
Zygoptera



Abundance



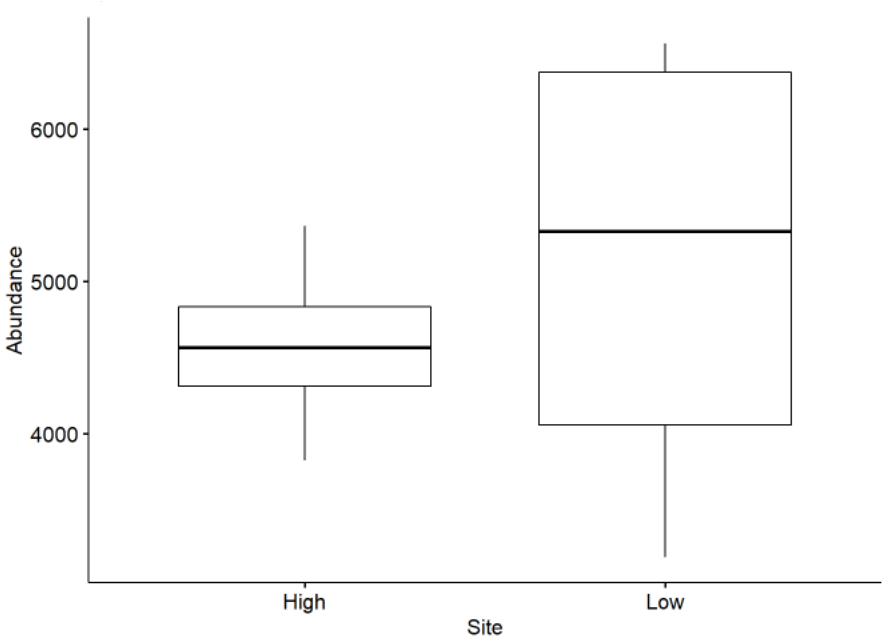

```
dat1$Family
```

```
##  [1] "Baetidae"              "Caenidae"              "Heptageniidae"
##  [4] "Perlidae"              "Hydropsychidae"        "Lepidostomatidae"
##  [7] "Ceratopogonidae"       "Chironomidae"          "Simuliidae"
## [10] "Tipulidae"             "Zygoptera"             "Elmidae"
## [13] "Helodidae"             "Gerridae"              "Potamonautidae"
## [16] "Sphaeriidae"           "Total abundance"       "Species richness"
## [19] "EPT Total"             "EPT%"                  "Species diversity (H)"
```

```
datab = dat1[dat1$Family == "Total abundance",]
datab1 = data.frame(t(datab)[-c(1:2),])
datab1$group = c(rep("Low",4), rep("High",4))
names(datab1) = c("Abundance", "Site")
datab1$Abundance = as.numeric(datab1$Abundance)
datab1$Site = as.factor(datab1$Site)

datab1 %>%
  group_by(Site) %>%
  get_summary_stats(Abundance, type = "mean_sd")
```

```
## # A tibble: 2 x 5
##   Site  variable      n  mean    sd
##   <fct> <chr>     <dbl> <dbl> <dbl>
## 1 High  Abundance     4 4579.  633.
## 2 Low   Abundance     4 5101. 1615.
```

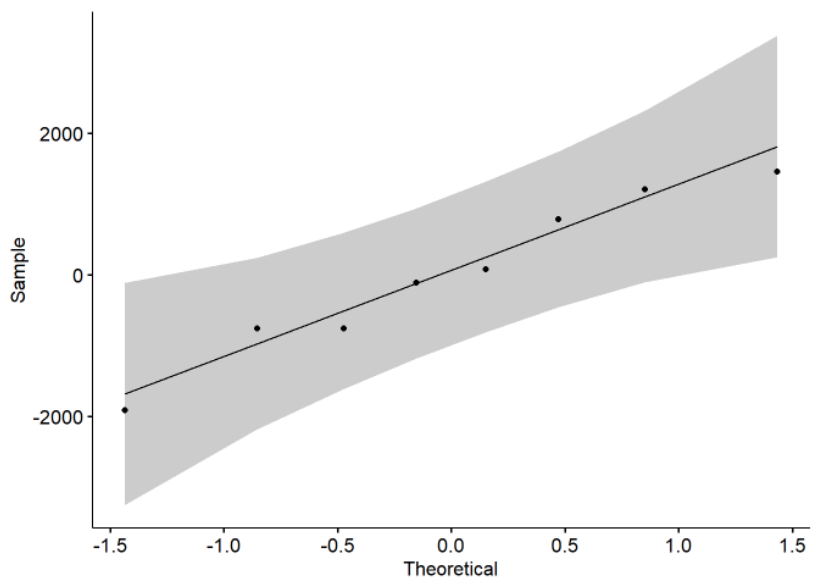
```
ggboxplot(datab1, x = "Site", y = "Abundance")
```




```
# Check outliers
datab1 %>%
  group_by(Site) %>%
  identify_outliers(Abundance)
```

```
## [1] Site       Abundance  is.outlier is.extreme
## <0 rows> (or 0-length row.names)
```

```
# Check normality
# Build the linear model
model  <- lm(Abundance ~ Site, data = datab1)
# Create a QQ plot of residuals
ggqqplot(residuals(model))
```



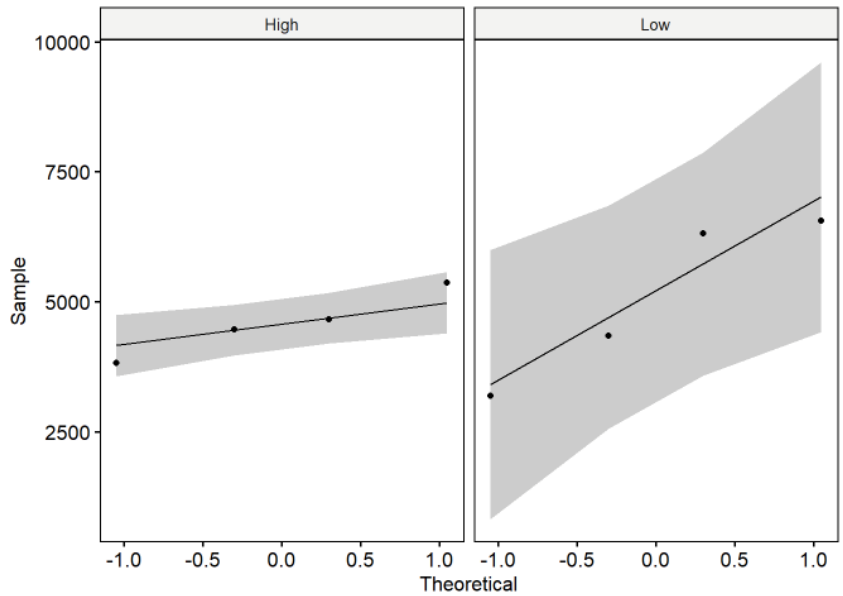
```
# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
```

```
## # A tibble: 1 x 3
##   variable          statistic p.value
##   <chr>                 <dbl>   <dbl>
## 1 residuals(model)      0.959   0.804
```

```
datab1 %>%
  group_by(Site) %>%
  shapiro_test(Abundance)
```

```
## # A tibble: 2 x 4
##   Site  variable  statistic     p
##   <fct> <chr>         <dbl> <dbl>
## 1 High  Abundance     0.986 0.936
## 2 Low   Abundance     0.894 0.404
```

```
ggqqplot(datab1, "Abundance", facet.by = "Site")
```



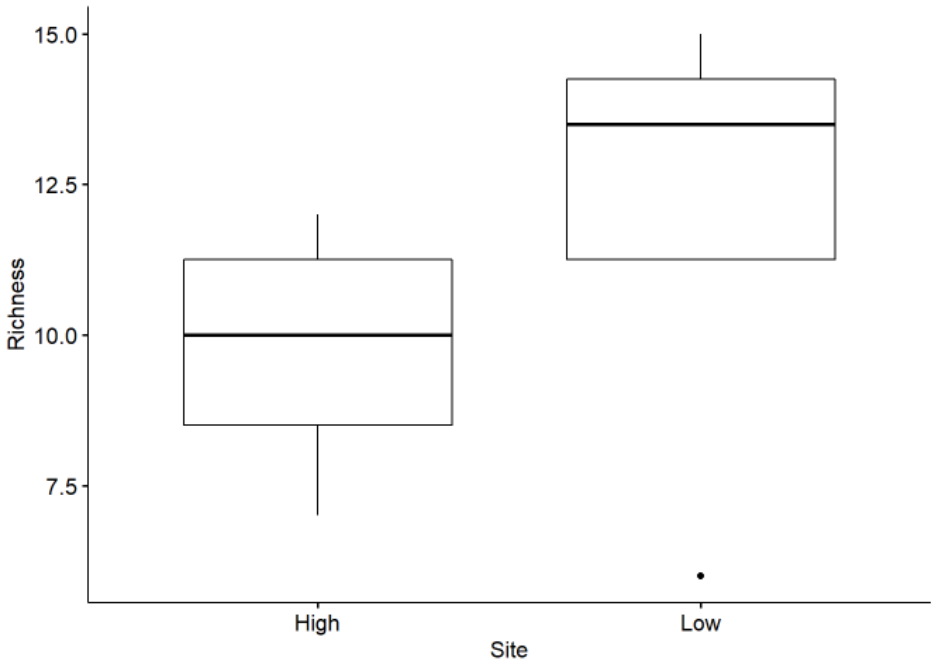

```
# t-test

t.test(Abundance ~ Site, data = datab1)
```

```
##
##  Welch Two Sample t-test
##
## data:  Abundance by Site
## t = -0.60214, df = 3.9009, p-value = 0.5803
## alternative hypothesis: true difference in means between group High and group Low is not equal to 0
## 95 percent confidence interval:
##  -2954.223  1909.873
## sample estimates:
## mean in group High  mean in group Low
##           4578.650           5100.825
```

```
# dat1$Family
datab = dat1[dat1$Family == "Species richness",]
datab1 = data.frame(t(datab)[-c(1:2),])
datab1$group = c(rep("Low",4), rep("High",4))
names(datab1) = c("Richness", "Site")
datab1$Richness = as.numeric(datab1$Richness)
datab1$Site = as.factor(datab1$Site)

datab1 %>%
  group_by(Site) %>%
  get_summary_stats(Richness, type = "mean_sd")
```

```
## # A tibble: 2 x 5
##   Site  variable     n  mean    sd
##   <fct> <chr>    <dbl> <dbl> <dbl>
## 1 High  Richness     4  9.75  2.22
## 2 Low   Richness     4 12     4.08
```

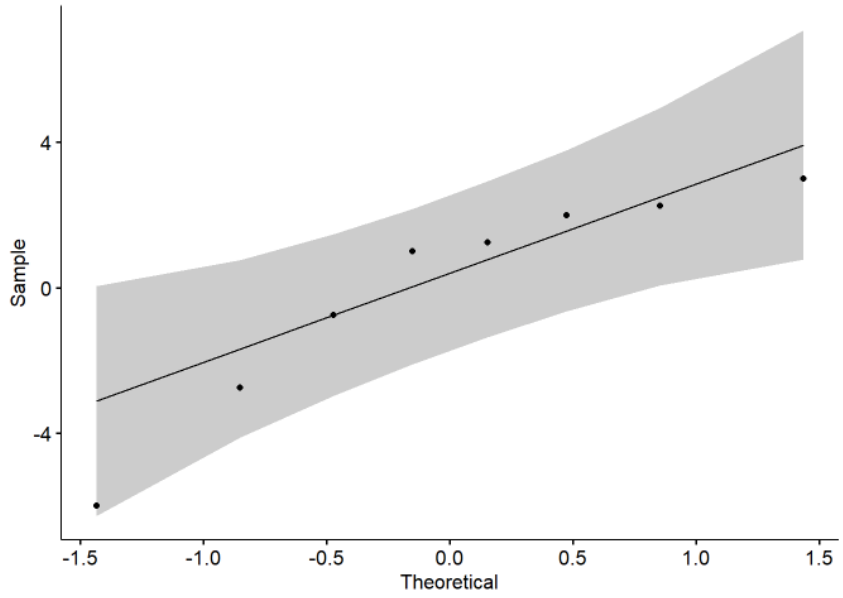
```
ggboxplot(datab1, x = "Site", y = "Richness")
```




```
# Check outliers
datab1 %>%
  group_by(Site) %>%
  identify_outliers(Richness)
```

```
## # A tibble: 1 x 4
##   Site  Richness is.outlier is.extreme
##   <fct>    <dbl> <lgl>      <lgl>
## 1 Low          6 TRUE       FALSE
```

```
# Check normality
# Build the linear model
model  <- lm(Richness ~ Site, data = datab1)
# Create a QQ plot of residuals
ggqqplot(residuals(model))
```



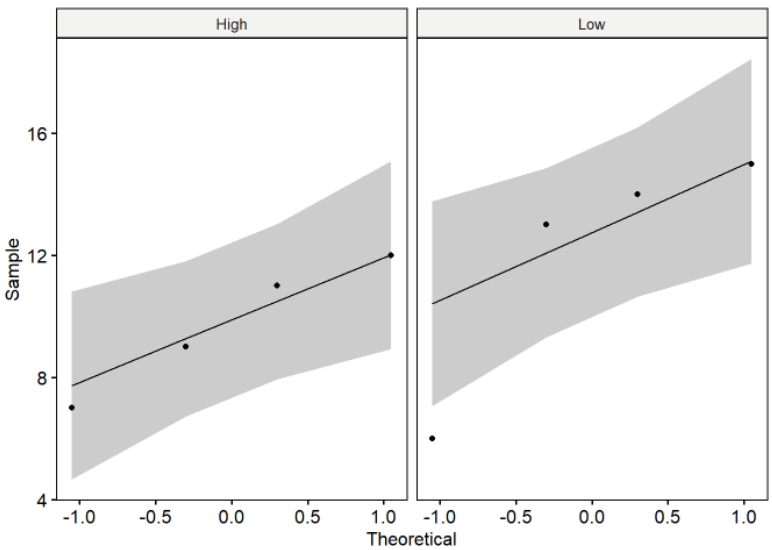

```
# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
```

```
## # A tibble: 1 x 3
##   variable         statistic p.value
##   <chr>                <dbl>   <dbl>
## 1 residuals(model)     0.874   0.165
```

```
datab1 %>%
  group_by(Site) %>%
  shapiro_test(Richness)
```

```
## # A tibble: 2 x 4
##   Site  variable statistic     p
##   <fct> <chr>        <dbl> <dbl>
## 1 High  Richness     0.963 0.798
## 2 Low   Richness     0.807 0.115
```

```
ggqqplot(datab1, "Richness", facet.by = "Site")
```



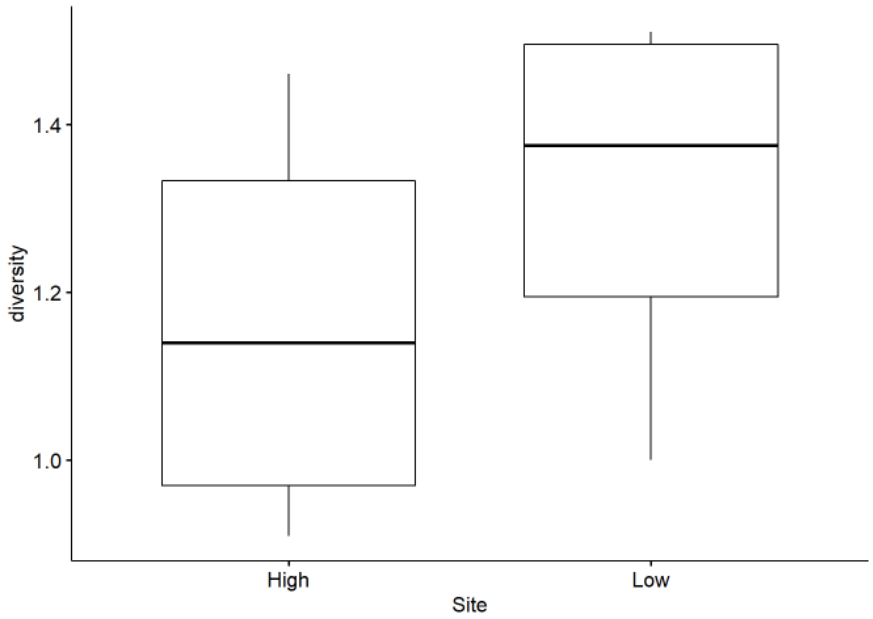
```r
# t-test

t.test(Richness ~ Site, data = datab1)
```

```
##
##  Welch Two Sample t-test
##
## data:  Richness by Site
## t = -0.96862, df = 4.6283, p-value = 0.3806
## alternative hypothesis: true difference in means between group High and group Low is not equal to 0
## 95 percent confidence interval:
##  -8.368349  3.868349
## sample estimates:
## mean in group High  mean in group Low
##               9.75              12.00
```

```r
# dat1$Family
datab = dat1[dat1$Family == "Species diversity (H)",]
datab1 = data.frame(t(datab)[-c(1:2),])
datab1$group = c(rep("Low",4), rep("High",4))
names(datab1) = c("diversity", "Site")
datab1$diversity = as.numeric(datab1$diversity)
datab1$Site = as.factor(datab1$Site)

datab1 %>%
  group_by(Site) %>%
  get_summary_stats(diversity, type = "mean_sd")
```

```
## # A tibble: 2 x 5
##   Site  variable      n  mean    sd
##   <fct> <chr>     <dbl> <dbl> <dbl>
## 1 High  diversity     4  1.16 0.257
## 2 Low   diversity     4  1.32 0.239
```

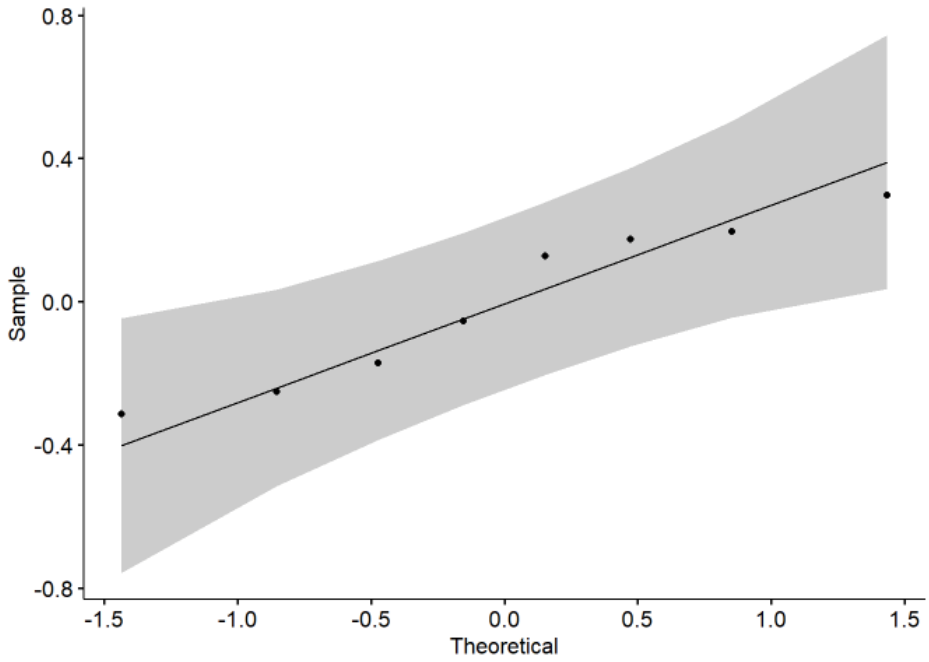
```r
ggboxplot(datab1, x = "Site", y = "diversity")
```



```
# Check outliers
datab1 %>%
  group_by(Site) %>%
  identify_outliers(diversity)
```

```
## [1] Site       diversity  is.outlier is.extreme
## <0 rows> (or 0-length row.names)
```

```
# Check normality
# Build the linear model
model  <- lm(diversity ~ Site, data = datab1)
# Create a QQ plot of residuals
ggqqplot(residuals(model))
```



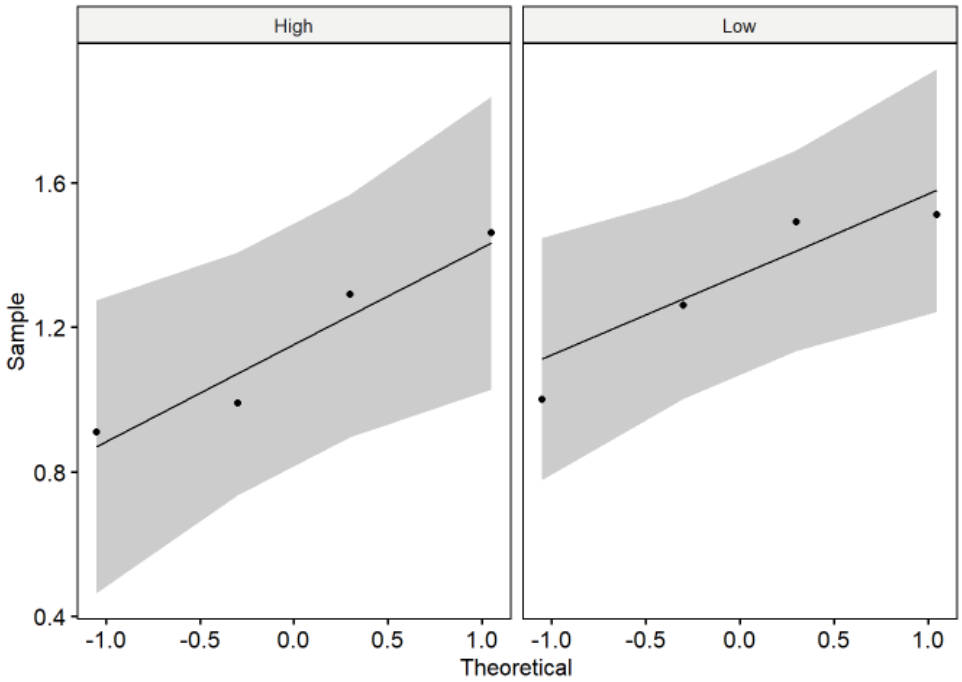

```
# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
```

```
## # A tibble: 1 x 3
##   variable         statistic p.value
##   <chr>                <dbl>   <dbl>
## 1 residuals(model)     0.922   0.443
```

```
datab1 %>%
  group_by(Site) %>%
  shapiro_test(diversity)
```

```
## # A tibble: 2 x 4
##   Site  variable  statistic     p
##   <fct> <chr>         <dbl> <dbl>
## 1 High  diversity     0.923 0.556
## 2 Low   diversity     0.884 0.358
```

```
ggqqplot(datab1, "diversity", facet.by = "Site")
```



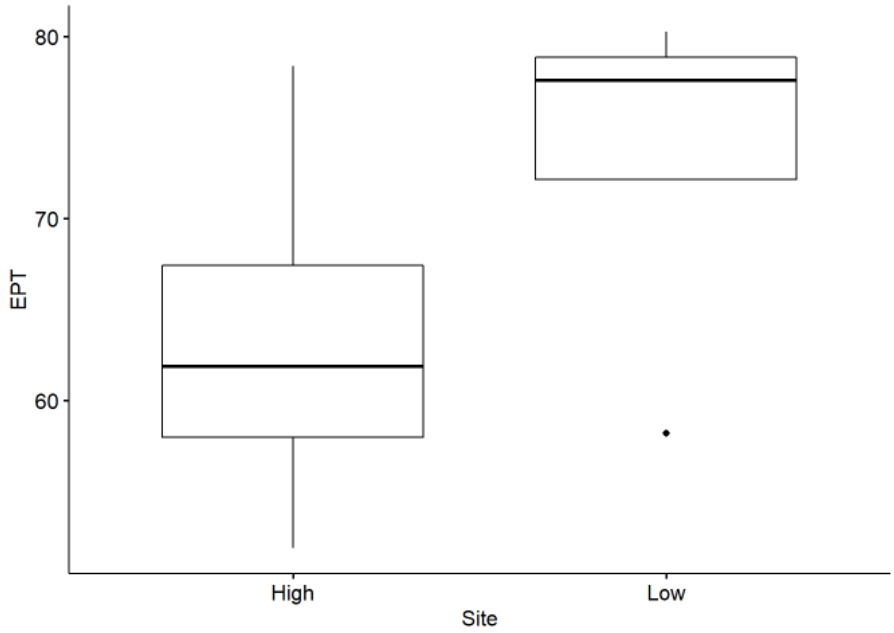

```
# t-test

t.test(diversity ~ Site, data = datab1)
```

```
##
##  Welch Two Sample t-test
##
## data:  diversity by Site
## t = -0.86945, df = 5.9672, p-value = 0.4182
## alternative hypothesis: true difference in means between group High and group Low is not equal to 0
## 95 percent confidence interval:
##  -0.582257  0.277257
## sample estimates:
## mean in group High  mean in group Low
##             1.1625             1.3150
```

```
# dat1$Family
datab = dat1[dat1$Family == "EPT%",]
datab1 = data.frame(t(datab)[-c(1:2),])
datab1$group = c(rep("Low",4), rep("High",4))
names(datab1) = c("EPT", "Site")
datab1$EPT = as.numeric(datab1$EPT)
datab1$Site = as.factor(datab1$Site)

datab1 %>%
  group_by(Site) %>%
  get_summary_stats(EPT, type = "mean_sd")
```

```
## # A tibble: 2 x 5
##   Site  variable     n  mean    sd
##   <fct> <chr>    <dbl> <dbl> <dbl>
## 1 High  EPT          4  63.5  11.1
## 2 Low   EPT          4  73.4  10.2
```

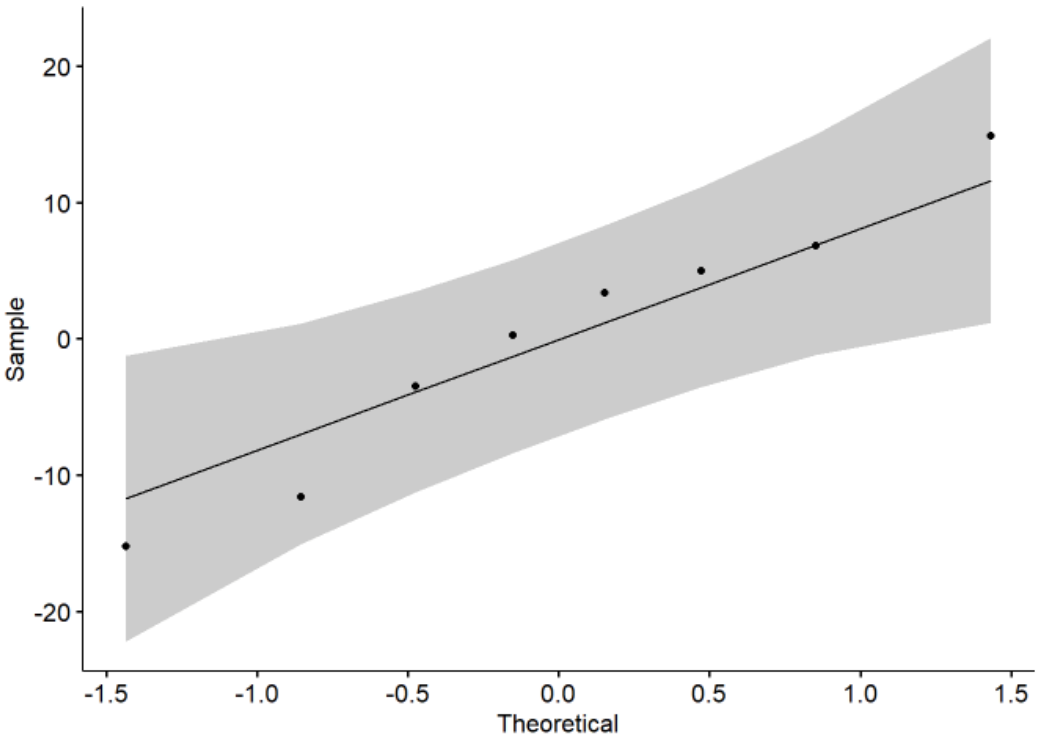
```
ggboxplot(datab1, x = "Site", y = "EPT")
```




```
# Check outliers
datab1 %>%
  group_by(Site) %>%
  identify_outliers(EPT)
```

```
## # A tibble: 1 x 4
##   Site    EPT is.outlier is.extreme
##   <fct> <dbl> <lgl>      <lgl>
## 1 Low    58.2 TRUE       FALSE
```

```
# Check normality
# Build the linear model
model  <- lm(EPT ~ Site, data = datab1)
# Create a QQ plot of residuals
ggqqplot(residuals(model))
```



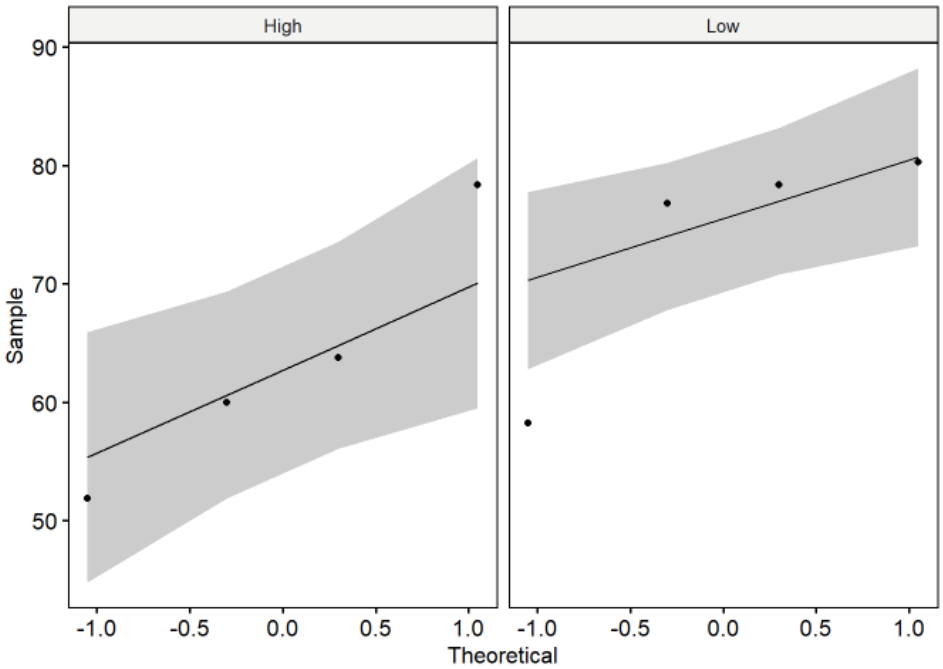
```
# Compute Shapiro-Wilk test of normality
shapiro_test(residuals(model))
```

```
## # A tibble: 1 x 3
##   variable         statistic p.value
##   <chr>                <dbl>   <dbl>
## 1 residuals(model)     0.967   0.876
```

```
datab1 %>%
  group_by(Site) %>%
  shapiro_test(EPT)
```

```
## # A tibble: 2 x 4
##   Site  variable statistic      p
##   <fct> <chr>        <dbl>  <dbl>
## 1 High  EPT          0.962 0.795
## 2 Low   EPT          0.756 0.0443
```

```
ggqqplot(datab1, "EPT", facet.by = "Site")
```



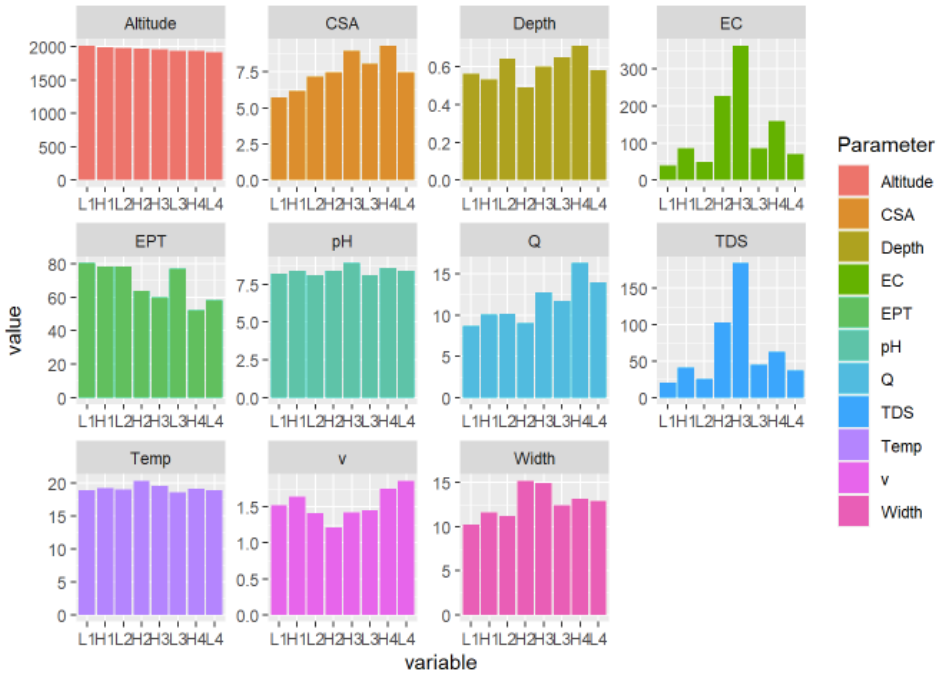

```
# t-test

t.test(EPT ~ Site, data = datab1)
```

```
##
##  Welch Two Sample t-test
##
## data:  EPT by Site
## t = -1.3126, df = 5.9622, p-value = 0.2376
## alternative hypothesis: true difference in means between group High and group Low is not equal to 0
## 95 percent confidence interval:
##  -28.374212   8.580812
## sample estimates:
## mean in group High  mean in group Low
##           63.51232           73.40902
```

```
dat2 = openxlsx::read.xlsx("Data/Cleaned_data_questions_Ibon.xlsx", 2)
dat2b = reshape2::melt(dat2)
```

```
## Using Parameter as id variables
```

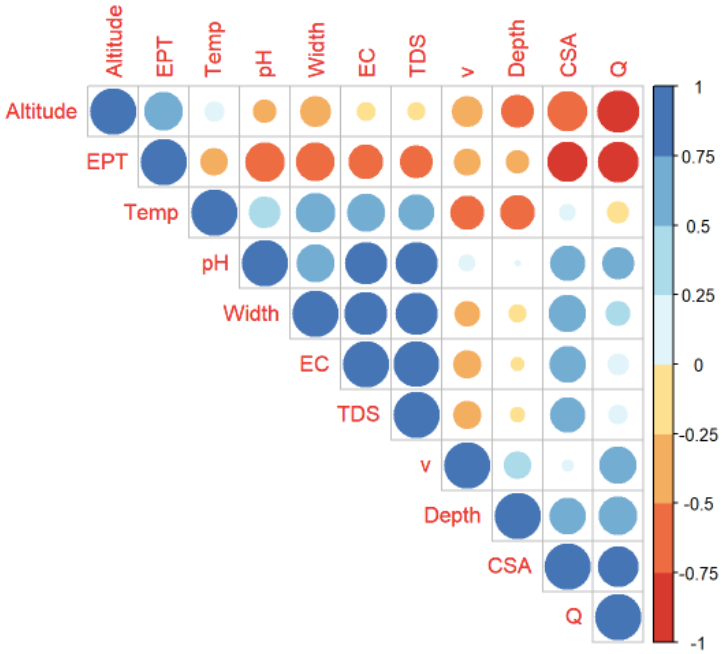
```
# ggplot(dat2b, aes(fill=Parameter, y=value, x=variable)) +
#     geom_bar(position="fill", stat="identity") + xlab("Site") + ylab("Abundance %")


ggplot(dat2b, aes(fill=Parameter, y=value, x=variable)) +
    geom_bar(position="dodge", stat="identity") +
    facet_wrap(~Parameter, scale="free")
```




```
library(corrplot)
```

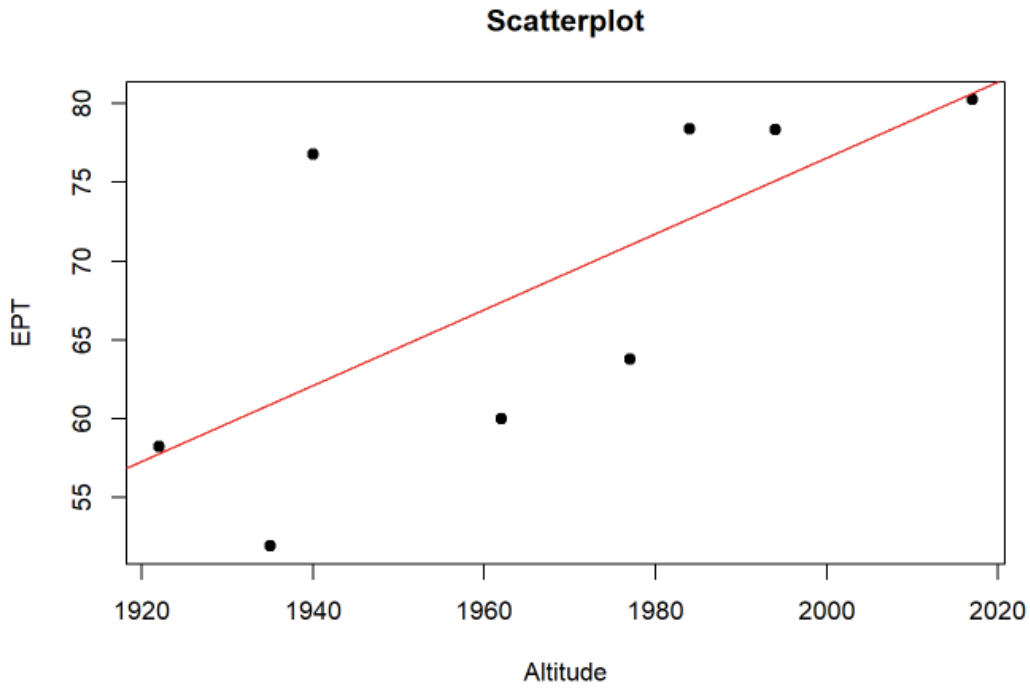
```
## corrplot 0.92 loaded
```

```
library(RColorBrewer)
dat2c = as.matrix(t(dat2[,-1]))
colnames(dat2c) = dat2$Parameter
M <-cor(dat2c)
corrplot(M, type="upper", order="hclust",
         col=brewer.pal(n=8, name="RdYlBu"))
```




```
##   Parameter          L1          H1          L2          H2         H3       L3
## 1  Altitude 2017.00000 1994.00000 1984.00000 1977.00000 1962.0000 1940.000
## 2     Width   10.20000   11.60000   11.20000   15.20000   14.9000   12.400
## 3     Depth    0.56000    0.53000    0.64000    0.49000    0.6000    0.650
## 4       CSA    5.71200    6.14800    7.16800    7.44800    8.9400    8.060
## 5         v    1.52000    1.64000    1.41000    1.21000    1.4200    1.450
## 6         Q    8.68224   10.08272   10.10688    9.01208   12.6948   11.687
##           H4          L4
## 1 1935.00000 1922.00000
## 2   13.10000   12.90000
## 3    0.71000    0.58000
## 4    9.30100    7.48200
## 5    1.75000    1.86000
## 6   16.27675   13.91652
```

```
dat2d = t(dat2[c(1,11),])
dat2e = data.frame(dat2d[-1,])
names(dat2e) = dat2d[1,]
dat2e$Site = rownames(dat2e)
dat2e$Site2 = gsub("1|2|3|4", "", dat2e$Site)
dat2e$Site3 = gsub("L|H", "", dat2e$Site)
dat2e$Site3 = as.factor(dat2e$Site3)
dat2e[,1] = as.numeric(dat2e[,1])
dat2e[,2] = as.numeric(dat2e[,2])

plot(dat2e[,1], dat2e[,2], main="Scatterplot",
   xlab=names(dat2e)[1], ylab=names(dat2e)[2], pch=19)
form = as.formula(paste0(names(dat2e)[2], "~",names(dat2e)[1]))
mod = lm(form, data = dat2e)
abline(mod, col="red")
```


Scatterplot
EPT
55
60
65
70
75
80
1920
1940
1960
1980
2000
2020
Altitude

```
summary(mod)
```

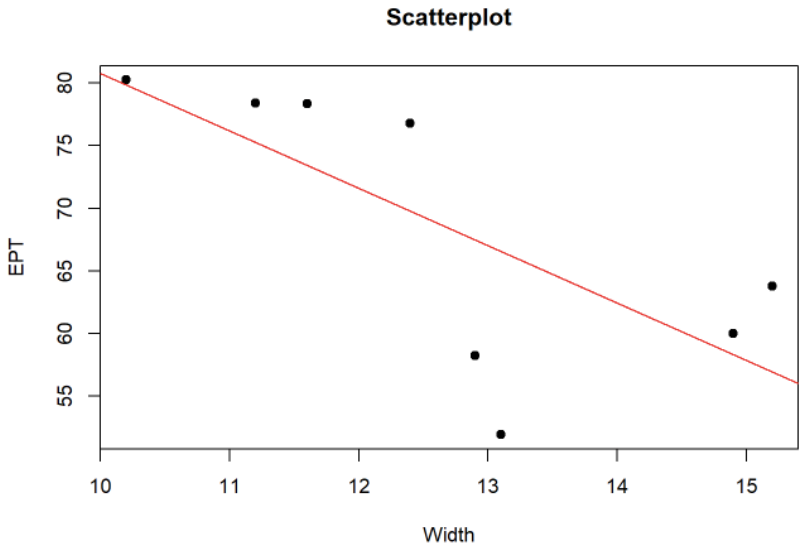
```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0036 -7.2845  0.0173  3.8612 14.6887
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -405.3981   197.1519  -2.056   0.0855 .
## Altitude       0.2410     0.1002   2.404   0.0530 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.634 on 6 degrees of freedom
## Multiple R-squared:  0.4906, Adjusted R-squared:  0.4057
## F-statistic: 5.778 on 1 and 6 DF,  p-value: 0.05302
```

```
form = as.formula(paste0(names(dat2e)[2], "~",names(dat2e)[1], "+",names(dat2e)[4]))
mod2 = lm(form, data = dat2e)
summary(mod2)
```

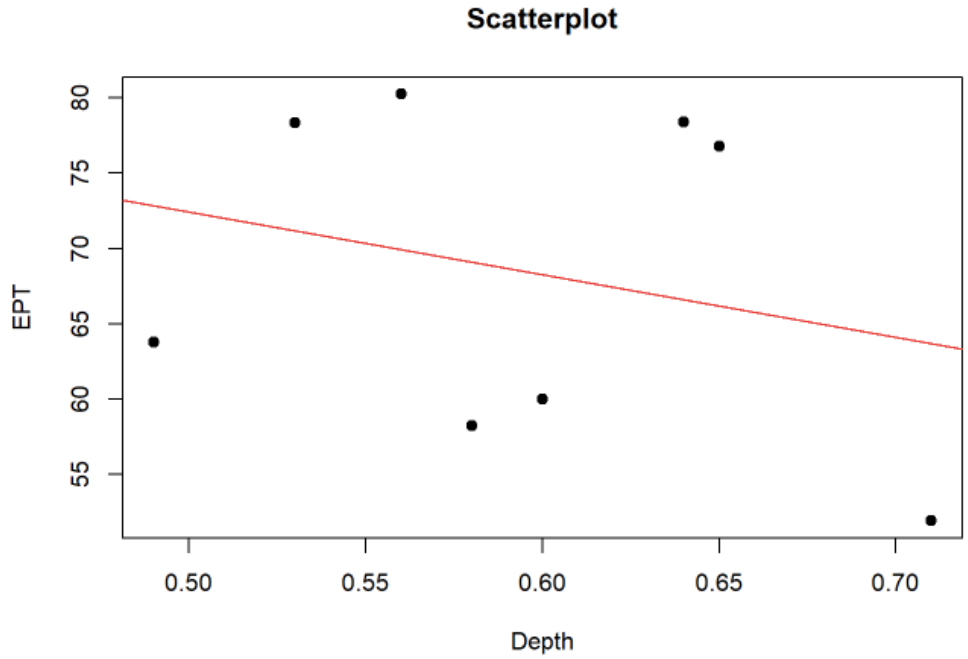
```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      L1      H1      L2      H2      H3      L3      H4      L4
## -5.6850  8.2636  0.5109 -2.1797 -2.2893  9.6783 -3.7946 -4.5041
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -417.26039  158.04051  -2.640   0.0460 *
## Altitude       0.24442    0.08033   3.043   0.0287 *
## Site2L        10.20222    4.89198   2.085   0.0914 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.917 on 5 degrees of freedom
## Multiple R-squared:  0.7276, Adjusted R-squared:  0.6186
## F-statistic: 6.677 on 2 and 5 DF,  p-value: 0.03874
```

```
# function

water_factors <- function(position, dat2)
{
    dat2d = t(dat2[c(position,11),])
      dat2e = data.frame(dat2d[-1,])
      names(dat2e) = dat2d[1,]
      dat2e$Site = rownames(dat2e)
      dat2e$Site2 = gsub("1|2|3|4", "", dat2e$Site)
      dat2e$Site3 = gsub("L|H", "", dat2e$Site)
      dat2e$Site3 = as.factor(dat2e$Site3)
      dat2e[,1] = as.numeric(dat2e[,1])
      dat2e[,2] = as.numeric(dat2e[,2])

      plot(dat2e[,1], dat2e[,2], main="Scatterplot",
         xlab=names(dat2e)[1], ylab=names(dat2e)[2], pch=19)
      form = as.formula(paste0(names(dat2e)[2], "~",names(dat2e)[1]))
      mod = lm(form, data = dat2e)
      abline(mod, col="red")
      print(summary(mod))

      form = as.formula(paste0(names(dat2e)[2], "~",names(dat2e)[1], "+",names(dat2e)[4]))
      mod2 = lm(form, data = dat2e)
      print(summary(mod2))
}
```

```
for(g in 2:10)
{
  water_factors(position = g, dat2)
}
```

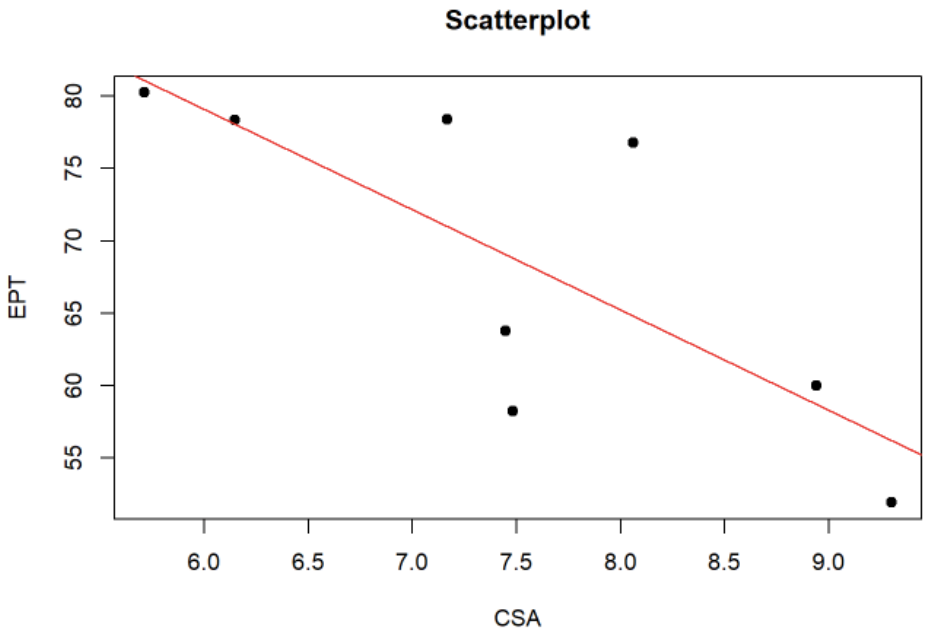



```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.672  -2.036   2.393   5.405   7.014
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  126.672     23.693   5.346  0.00175 **
## Width         -4.588      1.852  -2.477  0.04800 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.506 on 6 degrees of freedom
## Multiple R-squared:  0.5056, Adjusted R-squared:  0.4232
## F-statistic: 6.135 on 1 and 6 DF,  p-value: 0.048
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##       L1       H1       L2       H2       H3       L3       H4       L4
##   0.3549   5.6277   2.8827   6.8611   1.7658   6.5728 -14.2546  -9.8103
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 123.7606    35.8146   3.456   0.0181 *
## Width        -4.3977     2.5920  -1.697   0.1505
## Site2L        0.9914     8.4169   0.118   0.9108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.305 on 5 degrees of freedom
## Multiple R-squared:  0.5069, Adjusted R-squared:  0.3097
## F-statistic:  2.57 on 2 and 5 DF,  p-value: 0.1707
```



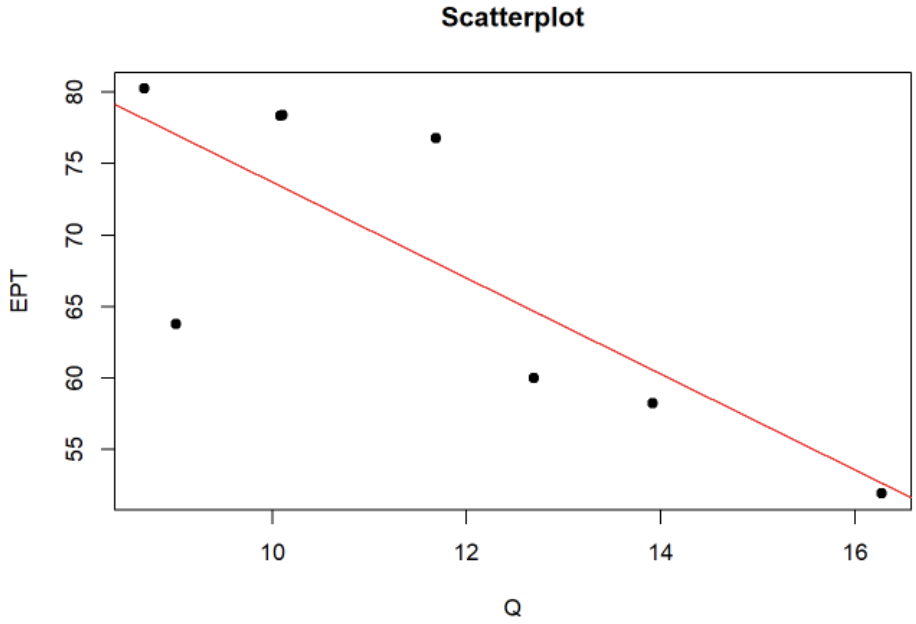

```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7868  -9.5025  -0.5189  10.4062  11.7894
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    93.18      37.36   2.494   0.0469 *
## Depth         -41.54      62.40  -0.666   0.5303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.67 on 6 degrees of freedom
## Multiple R-squared:  0.0688, Adjusted R-squared:  -0.0864
## F-statistic: 0.4433 on 1 and 6 DF,  p-value: 0.5303
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      L1      H1      L2      H2      H3      L3      H4      L4
##   4.099  11.831   6.848  -5.077  -2.501   5.839  -4.254 -16.785
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   97.149     34.262   2.835   0.0364 *
## Depth        -57.745     58.100  -0.994   0.3659
## Site2L        11.340      7.686   1.475   0.2001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.67 on 5 degrees of freedom
## Multiple R-squared:  0.3513, Adjusted R-squared:  0.09176
## F-statistic: 1.354 on 2 and 5 DF,  p-value: 0.339
```

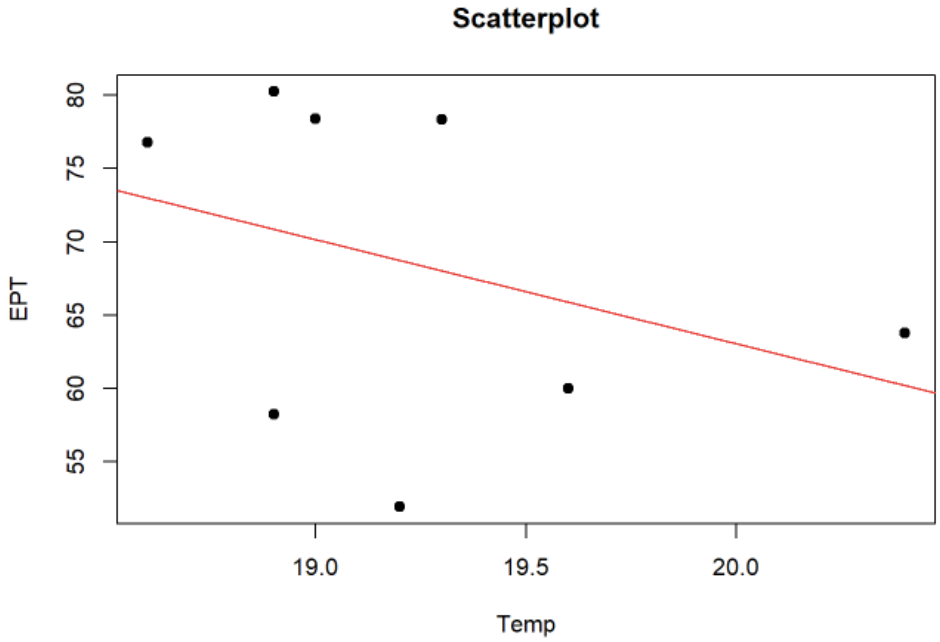



```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.652 -10.483   1.245   8.679  11.848
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    96.02      31.90   3.010   0.0237 *
## v             -17.99      20.65  -0.871   0.4172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.4 on 6 degrees of freedom
## Multiple R-squared:  0.1123, Adjusted R-squared:  -0.03568
## F-statistic: 0.7589 on 1 and 6 DF,  p-value: 0.4172
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     L1     H1     L2     H2     H3     L3     H4     L4
##  5.962 17.832  1.672 -6.224 -5.381  0.965 -6.227 -8.599
##
```

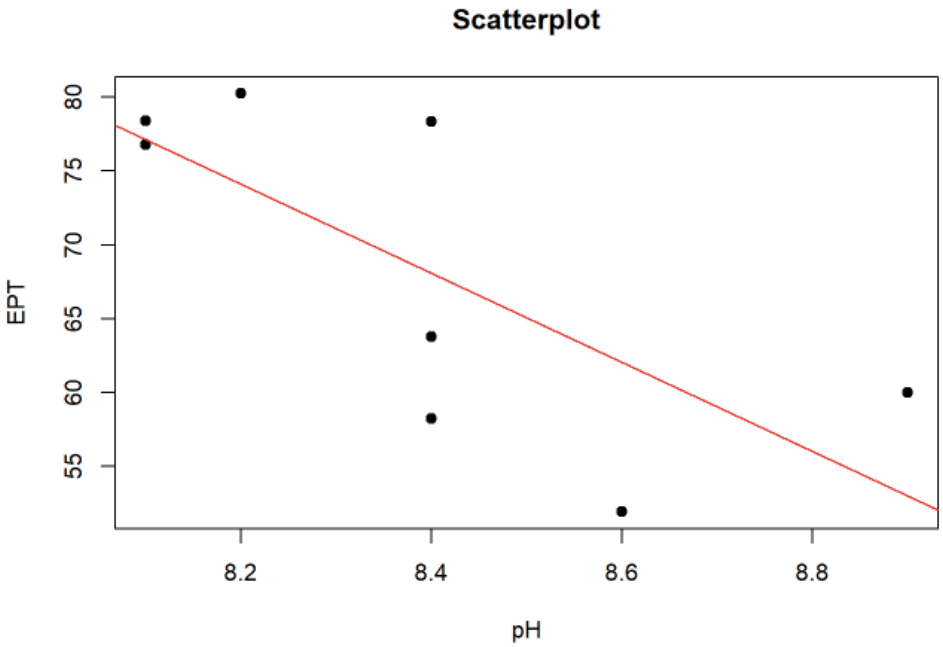
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   96.615     29.023   3.329   0.0208 *
## v            -21.995     18.974  -1.159   0.2987
## Site2L        11.106      7.406   1.500   0.1940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.37 on 5 degrees of freedom
## Multiple R-squared:  0.3877, Adjusted R-squared:  0.1427
## F-statistic: 1.583 on 2 and 5 DF,  p-value: 0.2934
```



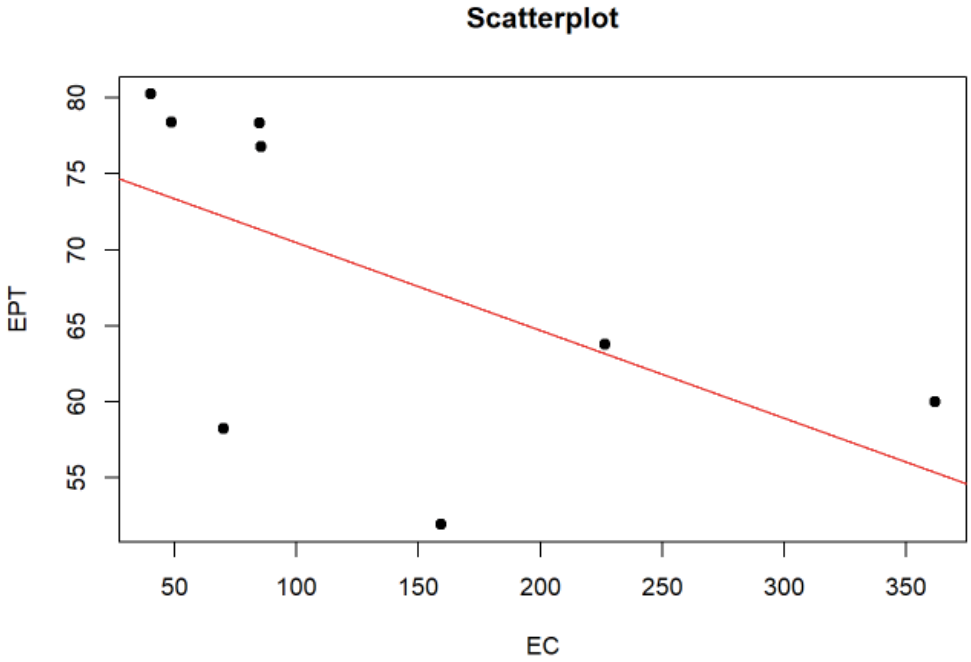

```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2371  -2.9004   0.7115   4.9807   8.7684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  107.299     12.702   8.447  0.00015 ***
## Q             -3.360      1.075  -3.126  0.02044 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.462 on 6 degrees of freedom
## Multiple R-squared:  0.6195, Adjusted R-squared:  0.5561
## F-statistic:  9.77 on 1 and 6 DF,  p-value: 0.02044
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     L1     H1     L2     H2     H3     L3     H4     L4
## -0.627  8.885  1.907 -9.023 -1.415  5.205  1.554 -6.485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  100.660     12.638   7.965 0.000503 ***
## Q             -3.091      1.012  -3.055 0.028255 *
## Site2L         7.058      4.966   1.421 0.214499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.899 on 5 degrees of freedom
## Multiple R-squared:  0.729,  Adjusted R-squared:  0.6206
## F-statistic: 6.725 on 2 and 5 DF,  p-value: 0.03823
```




```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.830  -7.577   3.687   8.525  10.358
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  204.957    147.543   1.389    0.214
## Temp          -7.095      7.667  -0.925    0.390
##
## Residual standard error: 11.32 on 6 degrees of freedom
## Multiple R-squared:  0.1249, Adjusted R-squared:  -0.02093
## F-statistic: 0.8565 on 1 and 6 DF,  p-value: 0.3904
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##       L1       H1       L2       H2       H3       L3       H4       L4
##   6.8475  14.8233   4.9898   0.3589  -3.5145   3.3540 -11.6677 -15.1914
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  65.9017   231.9252   0.284    0.788
## Temp         -0.1218    11.8141  -0.010    0.992
## Site2L        9.8023    12.3307   0.795    0.463
##
## Residual standard error: 11.68 on 5 degrees of freedom
## Multiple R-squared:  0.2231, Adjusted R-squared:  -0.08765
## F-statistic: 0.718 on 2 and 5 DF,  p-value: 0.532
```

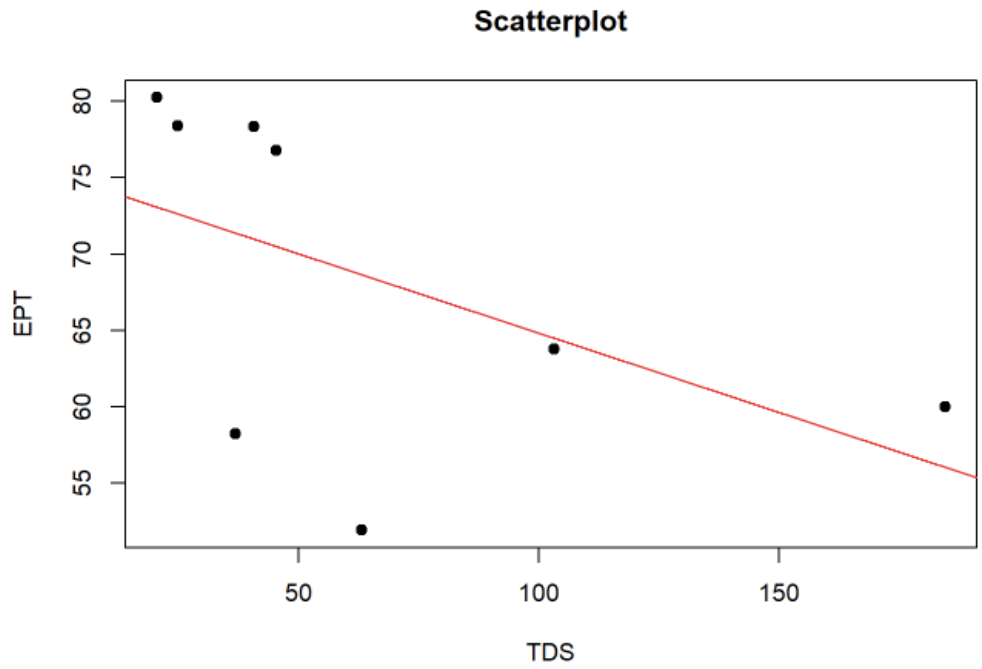



```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1417  -5.6973   0.4369   6.3496  10.2923
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   321.97      97.64   3.297   0.0165 *
## pH            -30.22      11.64  -2.597   0.0408 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.3 on 6 degrees of freedom
## Multiple R-squared:  0.5293, Adjusted R-squared:  0.4508
## F-statistic: 6.746 on 1 and 6 DF,  p-value: 0.04081
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##       L1       H1       L2       H2       H3       L3       H4       L4
##   6.8414   8.7444   1.4753  -5.8539   7.8515  -0.1118 -10.7419  -8.2049
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  363.318    161.682   2.247   0.0745 .
## pH           -34.963     18.848  -1.855   0.1228
## Site2L        -3.214      9.506  -0.338   0.7490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.99 on 5 degrees of freedom
## Multiple R-squared:  0.5398, Adjusted R-squared:  0.3557
## F-statistic: 2.932 on 2 and 5 DF,  p-value: 0.1437
```




```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.139  -3.027   4.812   5.703   7.038
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 76.23384    5.74917  13.260 1.14e-05 ***
## EC          -0.05774    0.03383  -1.707    0.139
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.925 on 6 degrees of freedom
## Multiple R-squared:  0.3269, Adjusted R-squared:  0.2147
## F-statistic: 2.913 on 1 and 6 DF,  p-value: 0.1387
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##      L1      H1      L2      H2      H3      L3      H4      L4
##   5.836   8.932   4.375   1.148   3.883   4.553 -13.964 -14.765
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 73.52078   12.12745   6.062  0.00176 **
## EC          -0.04809    0.05218  -0.922  0.39899
## Site2L       2.82646   10.82392   0.261  0.80441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.8 on 5 degrees of freedom
## Multiple R-squared:  0.3359, Adjusted R-squared:  0.07028
## F-statistic: 1.265 on 2 and 5 DF,  p-value: 0.3594
```




```
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.745  -3.818   4.879   6.508   7.402
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 75.21872    5.92450  12.696 1.46e-05 ***
## TDS         -0.10407    0.07153  -1.455    0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 6 degrees of freedom
## Multiple R-squared:  0.2608, Adjusted R-squared:  0.1376
## F-statistic: 2.117 on 1 and 6 DF,  p-value: 0.1959
##
##
## Call:
## lm(formula = form, data = dat2e)
##
## Residuals:
##       L1       H1       L2       H2       H3       L3       H4       L4
##   6.0217  10.7515   4.4682   0.6403   2.7249   4.3480 -14.1166 -14.8379
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 70.55760   11.24734   6.273  0.00151 **
## TDS         -0.07191    0.09980  -0.721  0.50346
## Site2L       5.14531   10.26074   0.501  0.63734
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.12 on 5 degrees of freedom
## Multiple R-squared:  0.2962, Adjusted R-squared:  0.01465
## F-statistic: 1.052 on 2 and 5 DF,  p-value: 0.4156
```

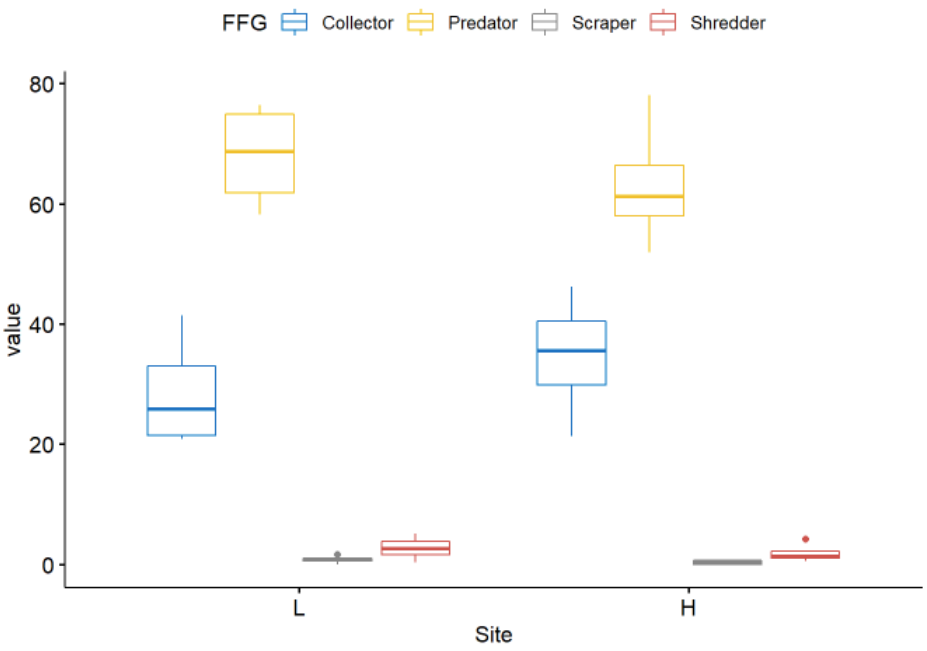
```
dat3 = openxlsx::read.xlsx("Data/Cleaned_data_questions_Ibon.xlsx", 3)
dat3b = dat3[,-c(6,11)]
dat3c = reshape2::melt(dat3b)
```

```
## Using FFG as id variables
```

```
dat3c$Site = gsub("1|2|3|4", "", dat3c$variable)

dat3c %>%
  group_by(Site, FFG) %>%
  get_summary_stats(value, type = "mean_sd")
```

```
## # A tibble: 8 x 6
##   FFG       Site  variable     n   mean     sd
##   <chr>     <chr> <chr>    <dbl>  <dbl>  <dbl>
## 1 Collector H     value        4 34.7   10.5
## 2 Predator  H     value        4 63.1   10.9
## 3 Scraper   H     value        4  0.339  0.392
## 4 Shredder  H     value        4  1.86   1.57
## 5 Collector L     value        4 28.5    9.60
## 6 Predator  L     value        4 68.0    8.80
## 7 Scraper   L     value        4  0.78   0.662
## 8 Shredder  L     value        4  2.70   2.04
```

```
bxp <- ggboxplot(
  dat3c, x = "Site", y = "value",
  color = "FFG", palette = "jco"
  )
bxp
```



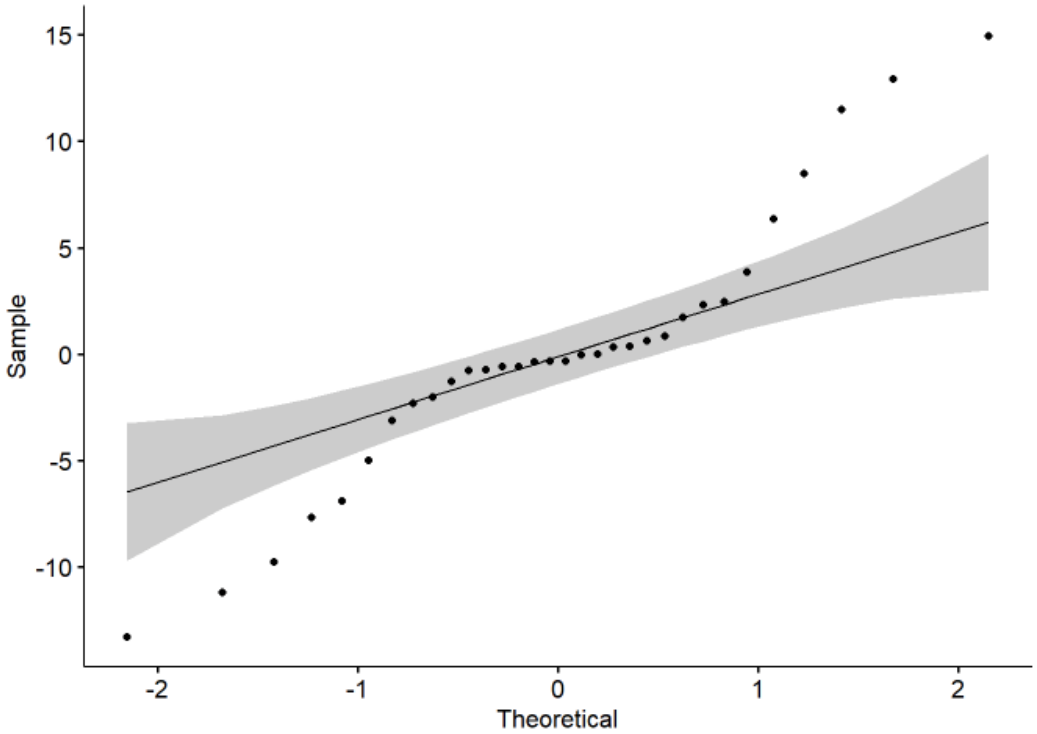
```
dat3c %>%
  group_by(Site, FFG) %>%
  identify_outliers(value)
```

```
## # A tibble: 2 x 6
##   FFG      Site  variable value is.outlier is.extreme
##   <chr>    <chr> <fct>    <dbl> <lgl>      <lgl>
## 1 Shredder H     H2        4.14 TRUE       FALSE
## 2 Scraper  L     L1        1.62 TRUE       FALSE
```

```
# Build the linear model
model  <- lm(value ~ Site*FFG,
             data = dat3c)
summary(model)
```

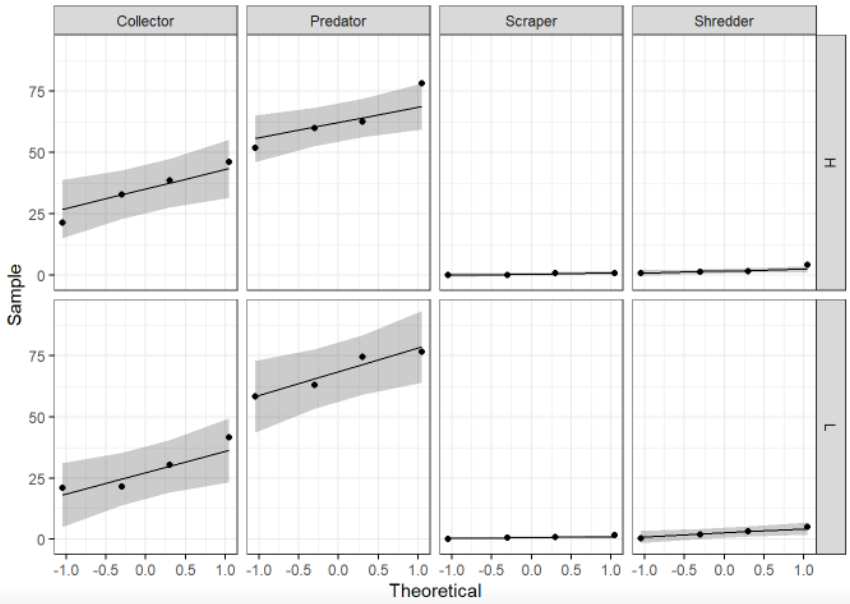
```
##
## Call:
## lm(formula = value ~ Site * FFG, data = dat3c)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.3091  -2.1201  -0.3395   1.8529  14.9454
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         34.677      3.563   9.732 8.33e-10 ***
## SiteL               -6.149      5.039  -1.220    0.234
## FFGPredator         28.451      5.039   5.646 8.18e-06 ***
## FFGScraper         -34.337      5.039  -6.814 4.77e-07 ***
## FFGShredder        -32.820      5.039  -6.513 9.78e-07 ***
## SiteL:FFGPredator   11.018      7.126   1.546    0.135
## SiteL:FFGScraper     6.590      7.126   0.925    0.364
## SiteL:FFGShredder    6.988      7.126   0.981    0.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.126 on 24 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9338
## F-statistic: 63.42 on 7 and 24 DF,  p-value: 6.173e-14
```

```
# Create a QQ plot of residuals
ggqqplot(residuals(model))
```




```
## # A tibble: 1 x 3
##   variable          statistic p.value
##   <chr>                 <dbl>   <dbl>
## 1 residuals(model)      0.934  0.0505
```

```
ggqqplot(dat3c, "value", ggtheme = theme_bw()) +
  facet_grid(Site ~ FFG)
```



```
res.aov <- dat3c %>% anova_test(value ~ Site * FFG)
```

```
## Coefficient covariances computed by hccm()
```

```
res.aov
```

```
## ANOVA Table (type II tests)
##
##     Effect DFn DFd       F        p p<.05   ges
## 1     Site   1  24   0.000 1.00e+00       0.000
## 2      FFG   3  24 147.161 1.39e-15     * 0.948
## 3 Site:FFG   3  24   0.819 4.96e-01       0.093
```