HW2: Statistics and Visualization

Exercise 1.1

Survey question used: How old are you?(Qn30)

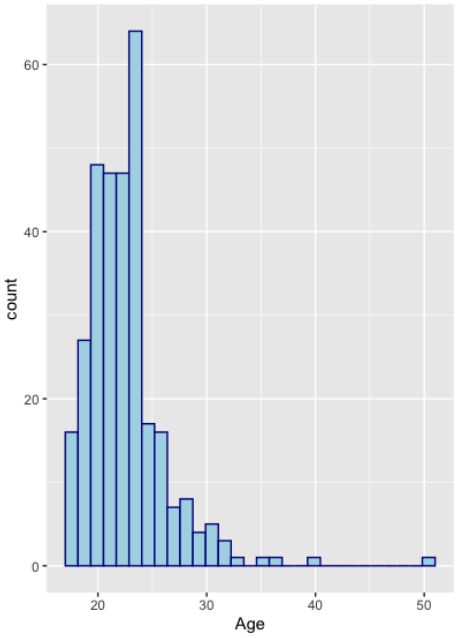

Figure 1: Histogram of age of survey respondents

Exercise 1.2

Survey question used: Do you identify as?(Qn29)

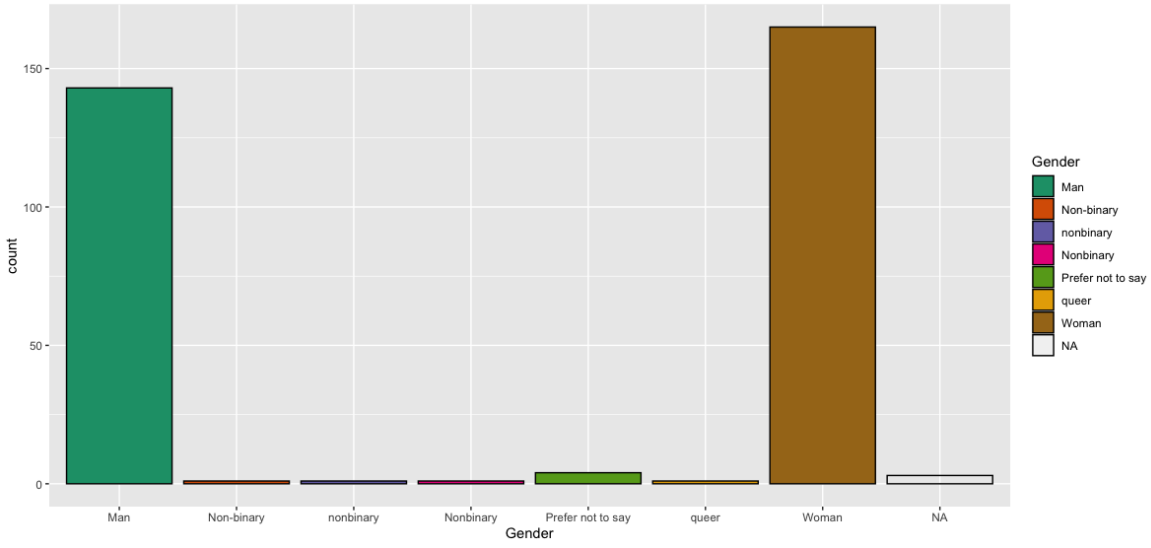

Figure 2: Barchart of genders of survey respondents

Exercise 1.3

Survey questions used: Do you rent or own your current living space? (Qn22) & How many miles do you live from the UW (Qn11)

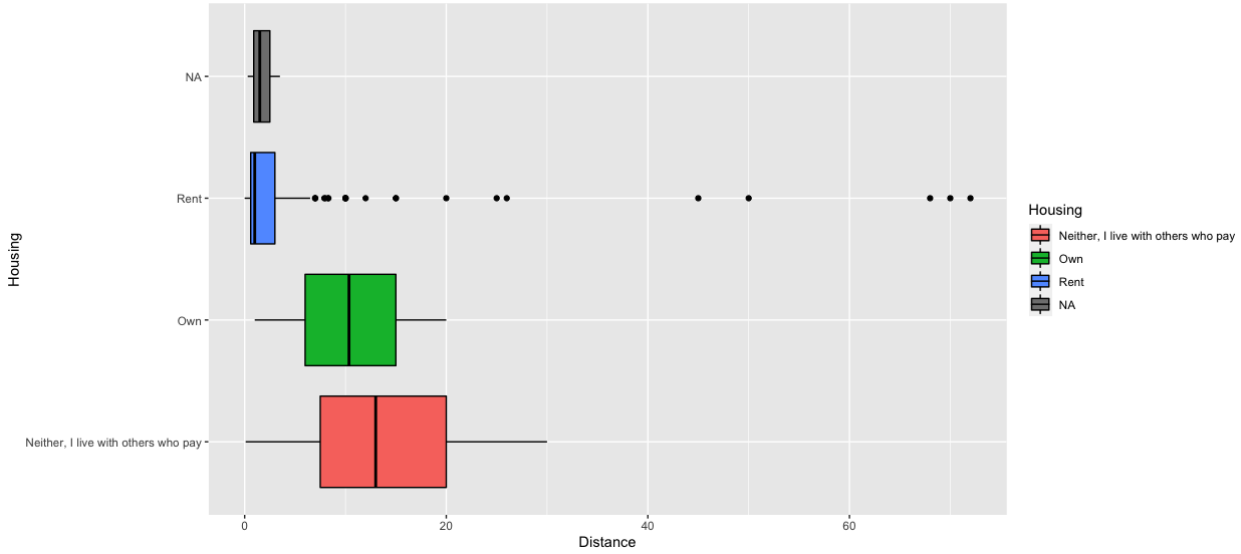

Figure 3: Boxplot of housing plotted against Distance from UW from survey respondents

Exercise 1.4

Survey questions used: Do you identify as?(Qn29) and How many miles do you live from the UW (Qn11)

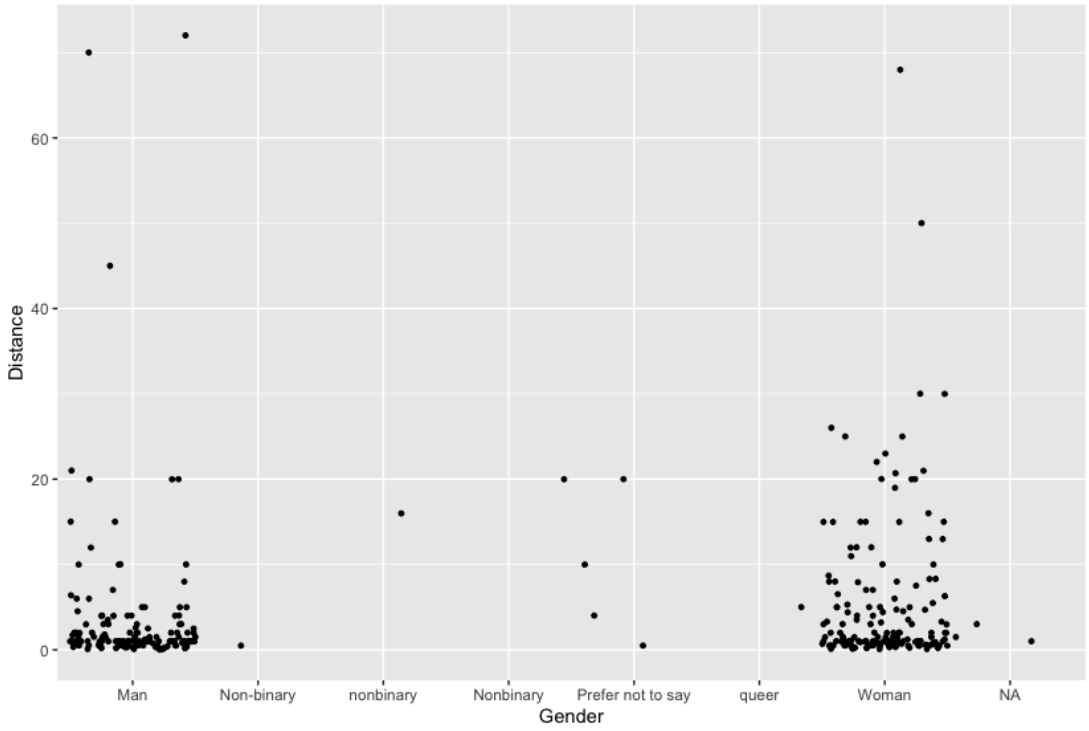

Figure 4: Scatterplot of Distance from UW plotted against Gender

Exercise 2: What is the difference between a histogram, barchart and a Pareto chart?
Histograms are used to indicate the distribution of non-discrete variables. Bar charts are used for comparison of discrete variables. A Pareto chart has similarities to histogram but displays the bars by the height of the bars, which signals the order of impact whereas histogram does not.

Exercise 3
A Welch Two Sample t-test was conducted against the two samples of How many miles do you live from the UW (Qn11) that are “Man” in Do you identify as?(Qn29) and, How many miles do you live from the UW (Qn11) that are “Woman” in Do you identify as?(Qn29). The Welch t-test was used because we assume that the samples are normally distributed and continuous with unequal variances.

Summary of How many miles do you live from the UW (Qn11) that are “Man” in Do you identify as?(Qn29)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.0	0.725	1.000	3.923	3.000	72.000	3

Summary of How many miles do you live from the UW (Qn11) that are “Man” in Do you identify as?(Qn29)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.10	0.70	1.50	5.50	6.45	68.00	4

After conducting the Welch Two Sample t-test, the p value obtained at the 95% confidence interval is 0.1421 which is more than 0.05. This implies that we reject the null hypothesis in which the two samples are the same and we do not reject the alternative hypothesis that the two samples are different.