STAT 260 R Assignment 1

Rodriguez Castro Raul V01030827

Question 1

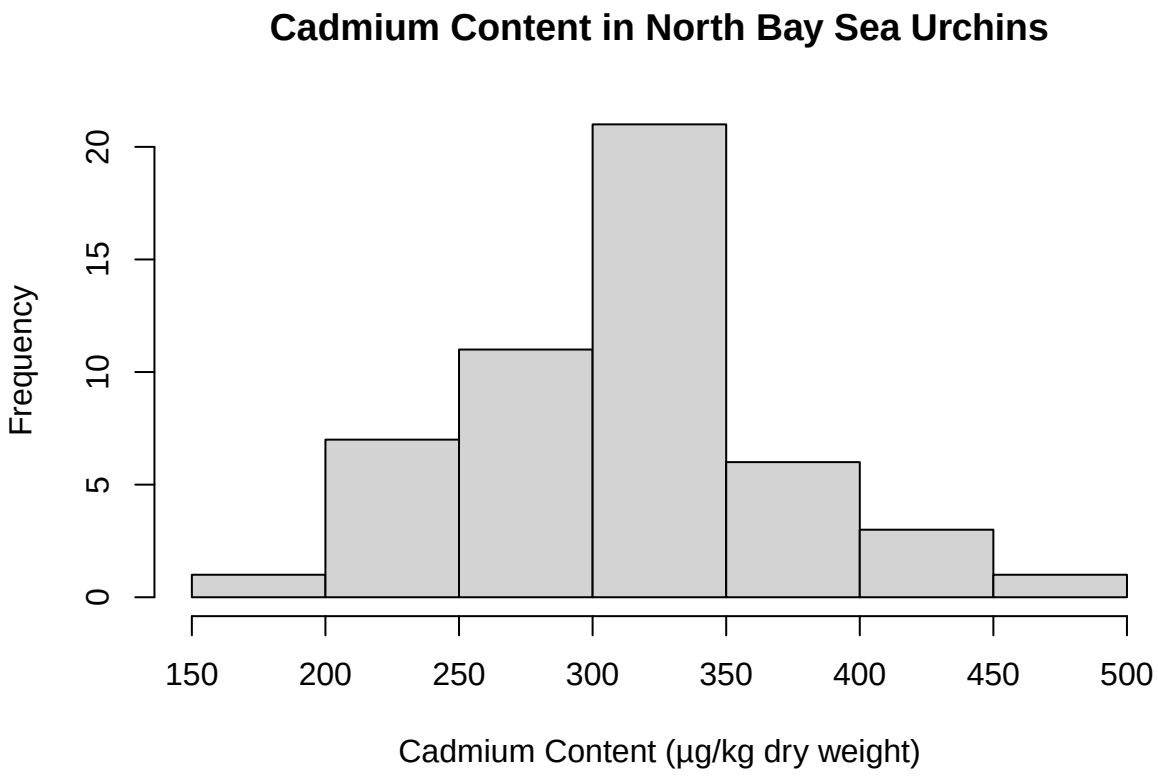
Part 1 Background: Echinoderms are considered useful bioindicators of water quality and marine sediments. If a region becomes polluted, sea urchins, with their minimal migration patterns, will be among the first species to show signs of the contamination. Scenario: Suppose that we want to compare the cadmium content of black sea urchins (Diadema antillarum) found in two neighbouring regions: North Bay and South Bay. Samples of adult urchins are collected from both regions, and the cadmium content of the urchins' internal tissues are measured as below, in µg of cadmium per kg of dry tissue weight. Cadmium in North Bay urchin tissue (µg/kg dry weight): 356 329 348 394 324 247 255 320 278 219 329 318 269 395 273 320 219 292 263 274 320 280 407 348 329 319 217 244 308 319 422 419 395 398 313 306 315 459 328 358 319 240 190 214 337 255 273 343 300 337 Cadmium in South Bay urchin tissue (µg/kg dry weight): 189 285 276 207 339 189 357 164 310 318 173 419 251 157 462 311 337 248 246 345 316 359 287 191 184 171 105 308 256 147 243 154 409 226 250 229 382 304 269 385 274 276 86 243 259

Note: These observations are also available in a text file on Crowdmark, if you wish to make use of the scan() function along with the copy and paste capabilities on your computer.

(a) [2 marks] Create a histogram of the cadmium content of the North Bay sea urchins. The title and x-axis for the histogram should have appropriate labels, including units where appropriate. Copy and paste the relevant command from the R Console Window and the resulting histogram into your Word document. (Do not include the code for how you stored the data into R, just include the one line of code for the histogram.)

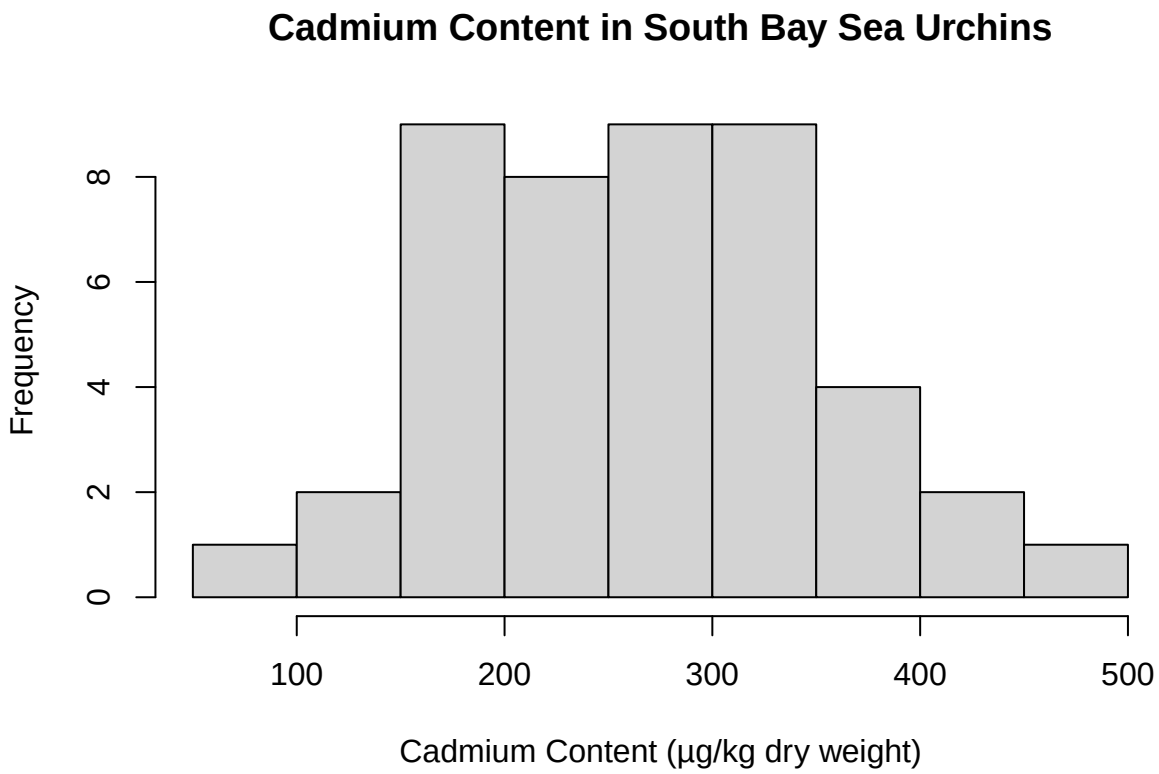
```
data <- scan("urchinData.txt", what=numeric(), skip=1, nlines=5)
north_bay <- data[1:50]
data <- scan("urchinData.txt", what=numeric(), skip=8)
south_bay <- data

hist(north_bay, main="Cadmium Content in North Bay Sea Urchins", xlab="Cadmium Content (µg/kg dry weigh
```



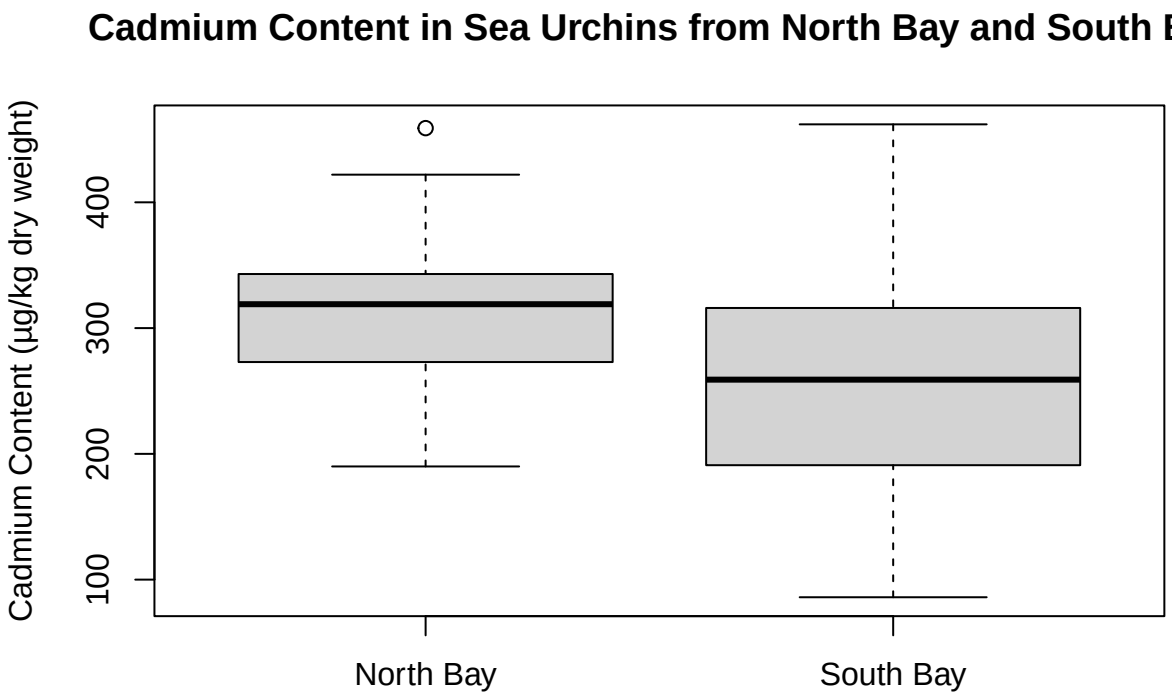
(b) [2 marks] Repeat the process in (a) for the South Bay urchins.

```
hist(south_bay, main="Cadmium Content in South Bay Sea Urchins", xlab="Cadmium Content (µg/kg dry weigh
```



(c) [2 marks] Create one side-by-side boxplot of the two sets of cadmium contents. The graph should be of a readable size, have an appropriate title, its x- and y-axes should have appropriate labels, and the two groups should be labelled. Copy and paste this boxplot and the line of code used to create it into your Word document.

```
boxplot(north_bay, south_bay, names=c("North Bay", "South Bay"), main="Cadmium Content in Sea Urchins fi
```



(d) [3 marks] Use R to calculate the means and standard deviations of the cadmium contents of the North Bay and South Bay urchins; for this question, use the dedicated commands for mean and standard deviation, not summary. Copy and paste the relevant commands and output from the R Console Window into your document. In plain English, write a short statement summarizing the values of your R output. Explicitly state the values and their units in your statement.

```
mean_north_bay <- mean(north_bay)
sd_north_bay <- sd(north_bay)
mean_north_bay
```

```
## [1] 312.68
```

```
sd_north_bay
```

```
## [1] 59.83291
```

```
mean_south_bay <- mean(south_bay)
sd_south_bay <- sd(south_bay)
mean_south_bay
```

```
## [1] 264.3556
```

```
sd_south_bay
```

```
## [1] 84.5862
```

(e) [2 marks] Answer the following question: Do North Bay sea urchins and South Bay urchins appear to have a different tissue cadmium content? If so, which population appears to have a greater cadmium content? Write a few plain English sentences explaining your opinion, making explicit references to the relevant features of the two data sets (including the mean, the median, the spread of the data, and the minimum and maximum values).

Use your results from both parts (c) and (d) to support your statement. You may wish to run the summary command in R for each sample to gather information about the maximum and minimum values (if you do so, include your copy/pasted R code here too).

The North Bay sea urchins appear to have a higher cadmium content in their tissues compared to the South Bay sea urchins. This is evident from the higher mean cadmium content (312.68 µg/kg in North Bay vs. 264.36 µg/kg in South Bay) and the more consistent levels of contamination as indicated by the lower standard deviation in North Bay (59.83 µg/kg) compared to South Bay (84.59 µg/kg). The boxplot also highlights that North Bay sea urchins have a generally higher range of cadmium content. Therefore, North Bay sea urchins appear to be more contaminated with cadmium than South Bay sea urchins.

Question 2

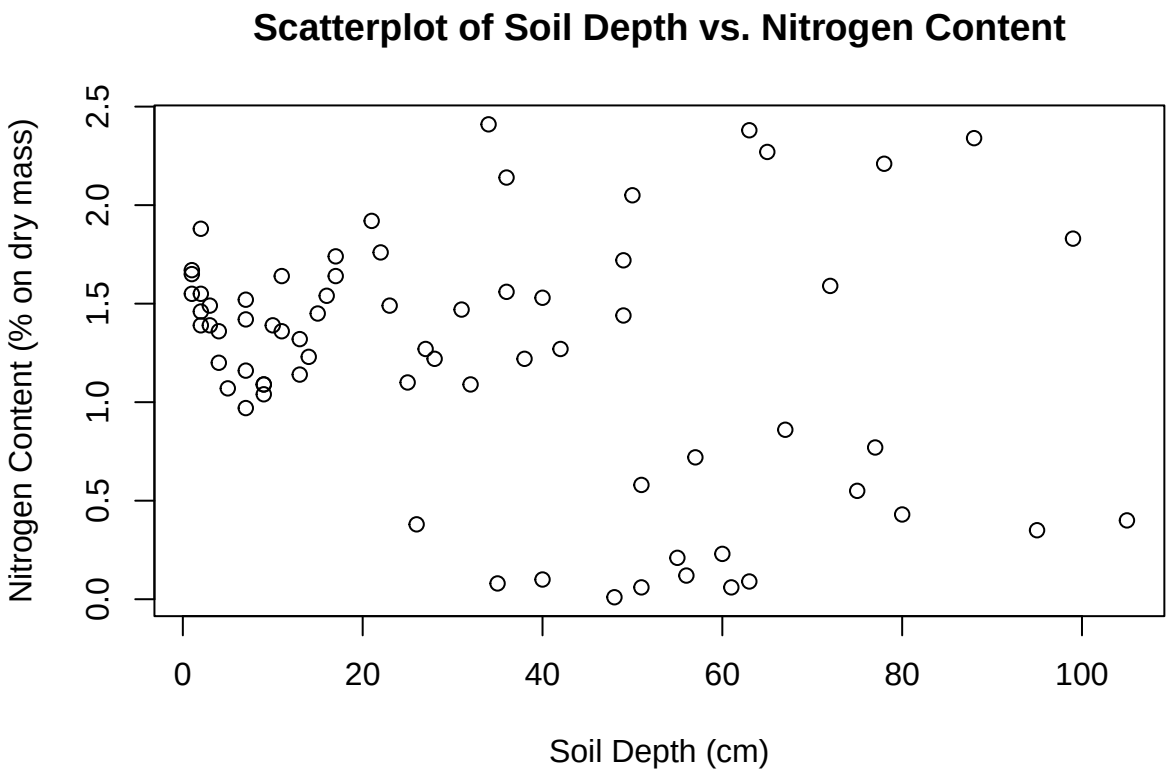
Part 2 Scenario: We will examine the relationship between soil depth (in cm) and nitrogen content (as percentage of dry mass) in permafrost samples taken from the Canadian subarctic. Use the bivariate dataset of 70 permafrost samples found in the permaFrost.csv1 file on Crowdmark.

(a) [2 marks] Create a scatterplot to compare the permafrost sample depth to nitrogen content. (Hint: here the depth is the value that can be controlled, so it is the x variable.) Your plot should have an appropriate title and the x- and y-axes should be labelled appropriately. Copy and paste the relevant commands and the plot into your Word document. (You do not need to include the code for how you stored the data into R, just include the line of code for the plot.)

```
permaFrost <- read.csv("permaFrost.csv")

x <- permaFrost$Soil_depth_cm
y <- permaFrost$Ncontent_percentage_on_drymass

plot(x,y,
     main="Scatterplot of Soil Depth vs. Nitrogen Content",
     xlab="Soil Depth (cm)",
     ylab="Nitrogen Content (% on dry mass)")
```



(b) [1 mark] In one or two sentences, describe the relationship (if any) between permafrost sample depth and nitrogen content that this scatterplot shows. (e.g. Is it linear or not? Positive or negative?)

The scatterplot shows a negative relationship between soil depth and nitrogen content. As soil depth increases, the nitrogen content tends to decrease. The relationship does not appear to be perfectly linear but shows a general trend of decreasing nitrogen content with increasing soil depth.

(c) [2 marks] Use the cor function to compute the correlation coefficient. (Copy and paste the relevant commands and output from the R Console Window into your Word document.) Does this value agree with your answer about the relationship permafrost sample depth and nitrogen content? Explain in a short sentence.

```
cov(x,y)
```

```
## [1] -4.386493
```

This value confirms the observed relationship in the scatterplot, where nitrogen content decreases as soil depth increases.