STATS 3042

Homework 1

Yevhen Myroniuk

A01005177 4r

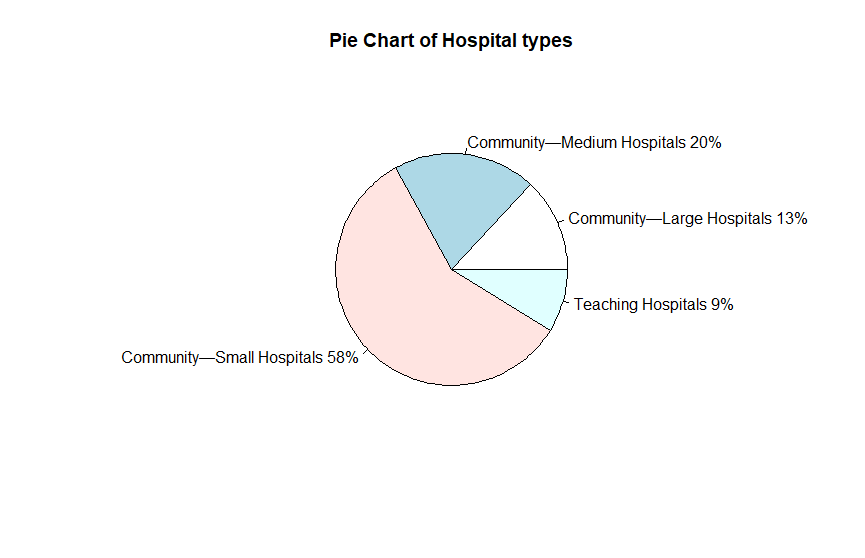
2018

**Homework 1 – Tables, Charts, and Graphs  
Due: September 28 as hard copy**

In this assignment, you will creating tables, charts, and graphs in order to make sense of the publicly available data about Canadian hospitals obtained from <https://open.canada.ca/data/en/dataset/ae1079f3-14fa-41dc-bde6-0b7c7f903c0e>. The data is available in the accompanying file hsp\_excerpt\_context\_en.xlsx, which you will need to import into R.

To import the Excel data, first save it to your local drive. Then, in RStudio, go **to File->Import Dataset->from Excel**. RStudio will tell you that you need to install a library. Accept this suggestion, and then wait until the next prompt. Enter the name of your file, and click **Import*.*** You will see this data in the top left window of RStudio, just like we viewed the built-in datasets. (If this doesn’t work, you may have to try again, possibly closing RStudio and reopening it.)

Note: since the field names contain spaces, R requires you to refer to them in single quotes, eg, ‘Type of Hospital’, which in turn complicates the use of certain commands. For instance, the second argument in the **filter** command will have to be **hsp\_excerpt\_context\_en$’Name of Column’**, not just **‘Name of Column’**. You may want to edit the headings in the Excel file before importing it to save yourself some typing/aggravation.

1. Create a pie chart of the hospitals by type. Display the percentages of each hospital type next to the category names. Make sure to give your chart an appropriate title.   
   Submit:  
   - The commands you used to create your pie chart, in order  
   - The pie chart  
   - An answer to this question: “What can we say about hospitals in Canada based on this pie chart?”  
     
   From inspecting the chart we can say that there are more small community hospitals than all others combined. Only about 1 out of 10 hospitals is teaching, all others are community hospitals.  
   

> hospital\_cats = aggregate(data.frame(count = hospital\_table$'Type of Hospital'), list(value=hospital\_table$'Type of Hospital'), length)

> hospital\_cats\_sum = sum(hospital\_cats$count)

> pct <- round(hospital\_cats$count/hospital\_cats\_sum\*100)

> lbls = c(hospital\_cats$value)

> lbls = paste(lbls, pct)

> lbls = paste(lbls, "%", sep="")

> pie(hospital\_cats$count, labels=lbls, main = "Pie Chart of Hospital types")

1. Create a scatterplot of Hospital Occupancy vs Number of Acute Care Beds. Be sure to filter out all the “NA” entries. Give your scatterplot an appropriate title and axis labels.

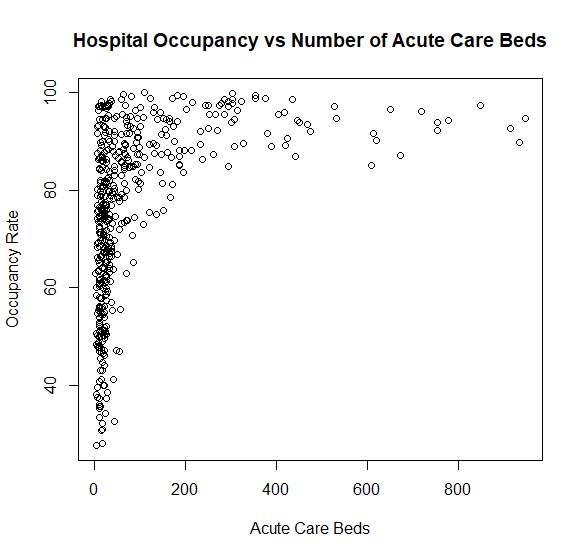
Submit:  
- The commands you used to create your scatterplot, in order  
- The scatterplot  
- An answer to this question: “What are two things your scatterplot tells you about the relationship between hospital occupancy and number of acute care beds?”

> nor = hospital\_table$`Hospital Occupancy Rate`[!is.na(hospital\_table$`Hospital Occupancy Rate`)]

> noa = hospital\_table$`Number of Acute Care Beds`[!is.na(hospital\_table$`Hospital Occupancy Rate`)]

|  |
| --- |
| > plot(nor~noa, ylab="Occupancy Rate", xlab="Acute Care Beds",  main="Hospital Occupancy vs Number of Acute Care Beds") |
|  |
| |  | | --- | |  | |

Two things we observe from the scatterplot are: most of the hospitals with low number of acute care beds have occupancy rate between 50 and 80 percent, and hospitals with high acute care bed number have high occupancy rate.



1. Find the mean number of acute care beds, by type of hospital. Be sure to use a data set that doesn’t contain any NA entries in the relevant fields.  
   Submit:  
   - The command(s) you used to obtain your data  
   - The four mean values computed

- An answer to this question: “Do the mean number of beds seem to correspond to hospital sizes? What do your numbers suggest about teaching hospitals in Canada?”

> f\_table = hospital\_table[complete.cases(hospital\_table),]

> aggregate(f\_table[, 6], list(f\_table$`Type of Hospital`), mean)

Group.1 Number of Acute Care Beds

1 Community—Large Hospitals 261.28846

2 Community—Medium Hospitals 78.89899

3 Community—Small Hospitals 19.50512

4 Teaching Hospitals 442.86111

The number of beds does correspond to hospital sizes. The number of beds is extremely high in teaching hospitals suggesting that medical education is highly prioritized in Canadian health system.

1. Create a stem plot of the number of acute care beds.   
   Submit:  
   - The command(s) you used to create your stem plot, in order  
   - An answer to these question: Describe the distribution of acute care beds in Canadian hospitals. Canada is a large country, with most of its population concentrated in a small number of cities. How do you expect a stem plot of acute care beds in, for example, Hong Kong hospitals to compare?

> stem(hospital\_table$`Number of Acute Care Beds`)

The stem plot is strongly skewed to the lower number of care beds. It does make sense since most of the population is concentrated in small cities which removes the need of big bed capacity hospitals. Considering the size of the country and the traveling distance it makes sense to put more small hospitals across the country rather than few big ones. In Hong Kong it would be the opposite – HK is a huge city in both population and size, but the demographics are much denser than in Canada. Therefore, it would be better to have few huge hospitals rather than smaller ones.

1. Create back-to-back stem plots that compare the hospital occupancy in the provinces with the two largest numbers of hospitals, Ontario and Alberta. Define variables in such a way as to give your output nice labels, and format your stem plots so they are not cluttered.

Submit:  
- The commands you used to create your back-to-back stem plots, in order  
- The back-to-back stem plots, copied directly from your output  
- An answer to this question: “How do the occupancies of Alberta hospitals compare to the occupancies of Ontario hospitals?”

> ontario = filter(f\_table, f\_table$`Province/Territory` == 'Ontario')

> alberta = filter(f\_table, f\_table$`Province/Territory` == 'Alberta')

> alberta = alberta$`Hospital Occupancy Rate`

> ontario = ontario$`Hospital Occupancy Rate`

> stem.leaf.backback(alberta, ontario, m=1)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1 | 2: represents 12, leaf unit: 1

alberta ontario

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2 55| 3 |07 2

7 88754| 4 |47 4

25 998887655322110000| 5 |1249 8

36 99877655432| 6 |123445567889 20

(21) 998888775544443321110| 7 |0011223345555677789999 42

33 999432221111| 8 |0012344444556666677777777788899 (31)

21 887766544444322111000| 9 |111222334444444556677777788888 31

| 10 |0 1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

n: 90 104

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

From inspecting the stem plot we can see that Alberta has less occupied hospitals compared to Ontario. The median of Ontario is in mid 80’s, where occupation median of Alberta is in 70’s. In conclusion, Ontario hospitals are more occupied than Alberta’s.

1. Create a histogram with at least ten **nonempty** bins of the number of acute care beds in small community hospitals in Saskatchewan. Choose your bin sizes appropriately and make sure your title and axis labels are appropriately descriptive. (NOTE: the Excel file uses the em-dash (–), which R can display but not read. Therefore, if you try to filter on the string “Community—Small Hospitals” – even if you copy-paste it directly from R! – you will get **zero** results. This is annoying and illogical; however, it is the sort of thing you can expect to run into in your work, and you will have to develop your own workarounds, so you might as well get used to it now. Hint: install the package **stringr** andlook up the **str\_detect** command. Then find the number of acute care beds in Saskatchewan, in hospital types that contain the substring ‘Small’.)

Submit:

- The commands you used to create your histogram, in order

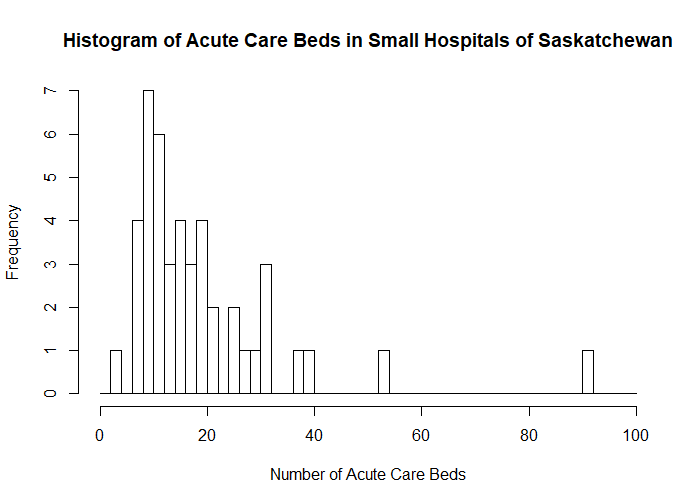
- The histogram

- An answer to this question: “Based on your histogram, a ‘typical’ small community hospital in Saskatchewan contains approximately how many acute care beds (you can give a range of values for your answer)? Are there any outliers among the small Saskatchewan hospitals? Explain.”

> cbs = filter(hospital\_table, hospital\_table$`Province/Territory`=='Saskatchewan' & hospital\_table$`Type of Hospital`=='Community—Small Hospitals' & !is.na(hospital\_table$`Number of Acute Care Beds`))

> breaks = seq(0, 100, by=2)

> hist(cbs$`Number of Acute Care Beds`, main="Histogram of Acute Care Beds in Small Hospitals of Saskatchewan", xlab="Number of Acute Care Beds", breaks=breaks)



Typical Hospital contains between 6 and 20 acute care beds. There is only 1 outlier that has 2 beds and 2 outliers for the bigger amount of beds. The reason behind it may be the fact that outlying hospitals with high number of acute care beds may be specializing on acute care beds only, being still a small hospital but with large amount of beds. Another possible explanation is the outdated data or incorrect evaluation of the hospital size.

**To submit: answers to all questions, including graphs, as a hard copy.**