Zuo Siqi 24445224
Stat-133 hw2 assignment

Read chapter 3 in book and do:
2-7, 2-8, 2-10, p 26
3-4, p 35
Read chapter 4 in book and do:
3-5, 3-6 p. 35
4-1, 4-5 p. 45

2-7
Problem 2.7
Install the nycflights13 package into R. (You can use the “Packages” tab which has an “install” button. If you are not using RStudio, given the R command install.packages("nycflights13")) Once the package is installed, you can access the flights data table with this command:
data(flights, package="nycflights13")
The codebook is available with help(flights) Using the codebook and examining the data table with the View() command (hint: you’ll need to give flights as an argument to View()), answer these questions:
1. How many variables are there?
2. How many cases are there?
3. What is the meaning of a case? (“Meaning” refers to the kind of entity, for instance, “airport” or “airline” or “date”. Hint: the case in flights is not any of these things.)
4. For each variable, is the variable quantitative or categorical?
5. For the variables air_time and distance, what are the units?

Ans:
1 ncol(flights) = 16 so 16 variables
2 nrow(flights) = 336776 so 336776 cases
3 it means a specific flight that has taken place
4 The variables:
year: qantitative
month: quantitative
day: quantitative
dep_time: quantitative
dep-delay: quantitative
arr_time: quantitative
arr_delay: quantitative
carrier: qualitative
tailnum: qualitative
flight: quantitative
origin: qualitative
dest: qualitative
airtime: quantitative
distance: quantitative
hour: quantitative

min: quantitative

5 minute and miles respectively

Problem 2.8
Consider this list of some possible mistakes in an assignment
operation:
1. No assignment operator
2. Unmatched quotes in character string
3. Improper syntax for function argument
4. Invalid object name
5. No mistake
For each of the following assignment statements, say
what is the mistake.

a. ralph <- sqrt 10
b. ralph2 <-- "Hello to you!"
c. 3ralph <- "Hello to you!"
d. ralph4 <- "Hello to you!
e. ralph5 <- date()

a Improper syntax for function argument
b Wrong assignment operator
c Invalid object name
d Unmatched quotes in character string
e No mistake

Problem 2.10
These questions should be easy to answer if you use the appropriate commands to load, view, or get documentation on the datasets.
• How many variables are there in CountryData?
• What does the variable tfat measure in the NCHS data table? (in package DataComputing)
• How many cases are there in WorldCities?
• What's the third variable in BabyNames?
• What are the codes for the levels of the categorical variable party in the RegisteredVoters data table, and what does each code stand for?

76 variables
tfat mass of trunk fat
23018 cases
count
DEM LIB REP UNA
DEM=Democratic, LIB=Libertarian,REP=Republican,UNA=Unaffiliated

Problem 3.4
Consider these R expressions. (You don't have to know
what the various functions do to solve this problem.)

```
Princes <-
        BabyNames %>%
        filter(name == "Prince") %>%
        group_by(year, sex) %>%
        summarise(yearlyTotal = sum(count))

# Now graph it!
Princes %>%
        ggplot(aes(x = year, y = yearlyTotal)) +
        geom_point(aes(color = sex)) +
        geom_vline(xintercept = 1978)
```

There are several kinds of named objects in the above expressions.
a. function name
b. data table name
c. variable name
d. name of a named argument

Using the naming convention and position rules, identify what kind of object each of the following name is used for. That is, assign one of the types (a) through (d) to each name.

1) BabyNames
data table name
2) filter
function name
3) name
variable name
4) ==
function name

2) group_by
function name
6) year
variable name
7) sex
variable name
8) summarise
function name

3) yearlyTotal
variable name
10) sum
function name
11) count
function name
12) ggplot
function name

4) aes

function name
14) x
name of a named argument
15) y
name of a named argument
16) geom_point
function name

5) color
name of a named argument
18) geom_vline
function name
19) xintercept
name of a named argument

Problem 3.5

There are several small, example data tables in the ggplot2 package. Look at the msleep data table by using the View() function with the name of the object as an argument.

- What is the meaning of the brainwt variable?
- How many cases are there?
- What is the real-world meaning of a case?
- What are the levels of the vore variable?

Ans:
Weight of brain of this species
83
Documentation and classification of a specific species
Carni, herbi, omni, insecti

Problem 3.6

The data verb functions all take a data table as their first argument and return a data table as their output. The chaining syntax lets the output of one function become the input to the following function, so you don't have to repeat the name of the data frame. An alternative syntax is to assign the output of one function to a named object, then use the object as the first argument to the next function in the computation.
Each of these statements, but one, will accomplish the same calculation. Identify the statement that does not match the others.

a) BabyNames %>%
group_by(year, sex) %>%
summarise(totalBirths=sum(count))
b) group_by(BabyNames, year, sex) %>%
summarise(totalBirths=sum(count))
c) group_by(BabyNames, year, sex) %>%

summarise(totalBirths=mean(count))
d) Tmp <- group_by(BabyNames, year, sex)
summarise(Tmp, totalBirths=sum(count))

Ans c

Problem 4.1

Markdown provides a simple way to produce section
headers and sub-headers, italic and bold text, monospaced
fonts suitable for computer commands, and even web
links. A reference is available in RStudio at the menu
Help/Markdown Quick Reference.
For each of the following, say how it will be rendered
when the Markdown is rendered to HTML. (Hint: You
can figure it out by reading the documentation, or you
can put the text into an Rmd document and compile it!)

one
italicised

two
bold

* three
bullet list form

Four
First level header

`five`
monospaced

Six
Second level header

[seven](http://tiny.cc/dcf/index.html)

Problem 4.5

From the RStudio console, load the DataComputing package like this: library(DataComputing) Once this is done, 1. Open a new file using the File/New File/R Markdown . . . menu item. Select “From Template” and then choose the DataComputing simple template. 2. Save the text that appears in the editor tab in a file named Birds.Rmd 3. Compile the Birds.Rmd file to HTML to verify that the template is working. 4. Edit the Birds.Rmd file to include contents that will make the compiled HTML file appear like this:

---
title: "Birds of the World"

````
author: "JJ Audubon"
date: ""
output:
  html_document:
    fig_height: 3
    fig_width: 5
---
<!-- Don't edit in between this line and the one below -->
```{r include=FALSE}
# Don't delete this chunk if you are using the DataComputing package
library(DataComputing)
```
*Source file*
```{r, results='asis', echo=FALSE}
includeSourceDocuments()
```
<!-- Don't edit the material above this line -->
````

There are many species of the birds in the world, from my studio I can see:

* Blue Jays

* Cardinals

* Robins

* Crows

* Sparrows