

HW2_Yun_Young

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():  dplyr, stats

##
## Please cite as:

## Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2. http://CRAN.R-project.org/package=stargazer

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last

## The following object is masked from 'package:purrr':
##
##   transpose

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##   hour, isoweek, mday, minute, month, quarter, second, wday,
##   week, yday, year

## The following object is masked from 'package:base':
##
##   date
```

Problem 1 Completed “R Programming E” lesson parts 4-7, 12 and“Getting and Cleaning Data” lessons parts 1-4. **Problem 2** Read through GitHub Chapters 1 and 2.Part A, B, C completed.
My repo set address on GitHub: https://github.com/stevewalsh124/STAT_5014.git

Problem 3 Created R Markdown file at the assigned location.

Problem 4 Version Control is a great back-up tool to store data as it captures different versions of the file. Having different files in development can allow you to experiment more freely and makes it easy to track down a mistake. Furthermore, users can easily compare different versions of files to track slight differences.

Problem 5 Sensory Data Analysis

Sensory data summary

Item	Person	value
Length:150	Length:150	Min. :0.700
Class :character	Class :character	1st Qu.:3.025
Mode :character	Mode :character	Median :4.700
NA	NA	Mean :4.657
NA	NA	3rd Qu.:6.000
NA	NA	Max. :9.400

Item	Person	value
------	--------	-------

Problem 5 LongJump Analysis

Long Jump data summary

YearCode	Year	dist
Min. :-4.00	Min. :1896	Min. :249.8
1st Qu.:21.00	1st Qu.:1921	1st Qu.:295.4
Median :50.00	Median :1950	Median :308.1
Mean :45.45	Mean :1945	Mean :310.3
3rd Qu.:71.00	3rd Qu.:1971	3rd Qu.:327.5
Max. :92.00	Max. :1992	Max. :350.5

Problem 5 BrainBody Weight

Brain/Body weight data summary

Brain	Body
Min. : 0.005	Min. : 0.10
1st Qu.: 0.600	1st Qu.: 4.25
Median : 3.342	Median : 17.25
Mean : 198.790	Mean : 283.13
3rd Qu.: 48.203	3rd Qu.: 166.00
Max. :6654.000	Max. :5712.00

Problem5 Tomato analysis

Tomato data summary

Variety	Clone	value
Length:18	Length:18	Length:18
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

Problem 6

```
##
## | Hi! I see that you have some variables saved in your workspace. To keep
## | things running smoothly, I recommend you clean up before starting swirl.
##
## | Type ls() to see a list of the variables in your workspace. Then, type
## | rm(list=ls()) to clear your workspace.
##
## | Type swirl() when you are ready to begin.
```

plants data summary

Accepted.Symbol	Scientific.Name	Foliage.Color	pH..Minimum.	pH..Maximum.	pH.Ratio
ABBA : 1	Abies balsamea : 1	Dark Green : 82	Min. :3.000	Min. : 5.100	Min. :0.4043
ACCA4 : 1	Acacia constricta : 1	Gray-Green : 25	1st Qu.:4.500	1st Qu.: 7.000	1st Qu.:0.6235
ACCO2 : 1	Acalypha virginica: 1	Green :692	Median :5.000	Median : 7.300	Median :0.6894
ACMI2 : 1	Acer negundo : 1	Red : 4	Mean :4.994	Mean : 7.345	Mean :0.6831
ACMIO : 1	Acer nigrum : 1	White-Gray : 9	3rd Qu.:5.500	3rd Qu.: 7.800	3rd Qu.:0.7429
ACNE2 : 1	Acer pensylvanicum: 1	Yellow-Green: 20	Max. :7.000	Max. :10.000	Max. :0.9355
(Other):826	(Other) :826	NA	NA	NA	NA

```
## Analysis of Variance Table
##
## Response: pH.Ratio
##           Df Sum Sq   Mean Sq F value Pr(>F)
## Foliage.Color   5 0.0772 0.0154346  1.8823  0.095 .
## Residuals    826 6.7730 0.0081998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 7

```
#This gives defect, and inspection details
Car_Gebreken_select <- fread(input = "Open_Data_RDW__Gebreken.csv", header = T, select=c(1,6), showProgress=F)

#This gives license plate, and inspection date, and defect code
Car_Geconstat_select <- fread(input = "Open_Data_RDW__Geconstateerde_Gebreken.csv", header=T, select=c(1,3,5),showProgress=F)

#this has the license plate, make and model of vehicle, and date.
Car_Person_select <- fread(input = "Personenauto_basisdata.csv", header=T, showProgress = F, select = c(1,3,4))

#selects items with 2017 in the specified column
Car_Geconstat_select_2017 <- Car_Geconstat_select[grep("2017",Car_Geconstat_select$"Meld datum door keuringsinstantie"),]
```