

COMPSCIX 415.2 Homework 5/Midterm

Santosh Kanutala

7/6/2018

Contents

Github location	1
RStudio and R Markdown	1
Question: 1	1
The tidyverse packages	2
Question: 1	2
Question: 2	2
R Basics	2
Question: 1	2
Question: 2	2
Question: 3	3
Data import/export	3
Question: 1	3
Question: 2	3
Visualization	4
Question: 1	4
Question: 2	4
Question: 3	4
Data munging and wrangling	4
Question: 1	4
Question: 2	4
Question: 3	4
EDA	4
Question: 1	4
Question: 2	4
Question: 3	4
Question: 4	4
Question: 5	4
Question: 6	4
Git and Github	4

Github location

My homework assignments can be found at <https://github.com/santumagic/compscix-415-2assignments.git>

RStudio and R Markdown

-

Question: 1

As part of this question, I have loaded the required packages and added instructions for table of contents etc in the YAML header.

```
# Load the required packages
library(tidyverse)
library(mdsr)
library(nycflights13)
```

The tidyverse packages

-

Question: 1

Plotting - **ggplot2**
 Data munging/wrangling - **dplyr** and **tidyr**
 Reshaping (speading and gathering) data - **tidyr**
 Importing/exporting data - **readr**

-

Question: 2

Plotting - **ggplot()** and **aes()**
 Data munging/wrangling - **select()** and **filter()**
 Reshaping (speading and gathering) data - **separate()** and **extract()**
 Importing/exporting data - **read_csv()** and **read_delim()**

R Basics

-

Question: 1

```
My_data.name___is.too00ooLong <- c( 1 , 2 , 3 )
```

Answer: Just with one change (removal of '!'), the code works.

-

Question: 2

```
# this is a character vector
my_string <- c('has', 'an', 'error', 'in', 'it')
my_string

## [1] "has"   "an"    "error" "in"    "it"
```

-

Question: 3

```
my_vector <- c(1, 2, '3', '4', 5)
my_vector
```

```
## [1] "1" "2" "3" "4" "5"
```

Answer: This is a numeric vector and with or without the single or double quotes, vector takes values.

Data import/export

-

Question: 1

```
# Download and import the file rail_train.txt
rail_train.txt <- read.delim("/Users/skanutal/Documents/Santosh/Learning/Berkeley/rail_train.txt", sep=
#glimpse the data from txt file
glimpse(rail_train.txt)
```

```
## Observations: 90
## Variables: 10
## $ hightemp    <int> 83, 73, 74, 95, 44, 69, 66, 66, 80, 79, 78, 65, 41,...
## $ lowtemp     <int> 50, 49, 52, 61, 52, 54, 39, 38, 55, 45, 55, 48, 49,...
## $ avgtemp     <dbl> 66.5, 61.0, 63.0, 78.0, 48.0, 61.5, 52.5, 52.0, 67....
## $ spring      <int> 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, ...
## $ summer      <int> 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, ...
## $ fall        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, ...
## $ cloudcover  <dbl> 7.6, 6.3, 7.5, 2.6, 10.0, 6.6, 2.4, 0.0, 3.8, 4.1, ...
## $ precip      <dbl> 0.00, 0.29, 0.32, 0.00, 0.14, 0.02, 0.00, 0.00, 0.0...
## $ volume      <int> 501, 419, 397, 385, 200, 375, 417, 629, 533, 547, 4...
## $ weekday     <int> 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, ...
```

-

Question: 2

```
# Export the .txt file as csv into a different location
rail_train_csv <- write_delim(
  rail_train.txt, delim = '|', path = "/Users/skanutal/Documents/Santosh/Learning/Berkeley/3. Intro to D
)
# Load the newly created csv file
rail_train_csv_final <- read_csv(
  "/Users/skanutal/Documents/Santosh/Learning/Berkeley/3. Intro to DS/Assignments/rail_train.csv", sep=
)
# glimpse the data from the final csv file
glimpse(rail_train_csv_final)
```

```
## Observations: 90
## Variables: 10
## $ hightemp    <int> 83, 73, 74, 95, 44, 69, 66, 66, 80, 79, 78, 65, 41,...
## $ lowtemp     <int> 50, 49, 52, 61, 52, 54, 39, 38, 55, 45, 55, 48, 49,...
```

```
## $ avgtemp    <dbl> 66.5, 61.0, 63.0, 78.0, 48.0, 61.5, 52.5, 52.0, 67....
## $ spring     <int> 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, ...
## $ summer     <int> 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, ...
## $ fall       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, ...
## $ cloudcover <dbl> 7.6, 6.3, 7.5, 2.6, 10.0, 6.6, 2.4, 0.0, 3.8, 4.1, ...
## $ precip     <dbl> 0.00, 0.29, 0.32, 0.00, 0.14, 0.02, 0.00, 0.00, 0.0...
## $ volume     <int> 501, 419, 397, 385, 200, 375, 417, 629, 533, 547, 4...
## $ weekday    <int> 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, ...
```

Visualization

-

Question: 1

-

Question: 2

-

Question: 3

Data munging and wrangling

Question: 1

Question: 2

Question: 3

EDA

Question: 1

Question: 2

Question: 3

Question: 4

Question: 5

Question: 6

Git and Github