MATH 3070 Lab Project 10

Pranav Rajan

November 5, 2020

Contents

Remember: I expect to see commentary either in the text, in the code with comments created using **#**, *or (preferably) both!* ***Failing to do so may result in lost points!***

Since this assignment involves simulation, I set the seed to the following in order to get the same results:

```
set.seed(5292016)
```

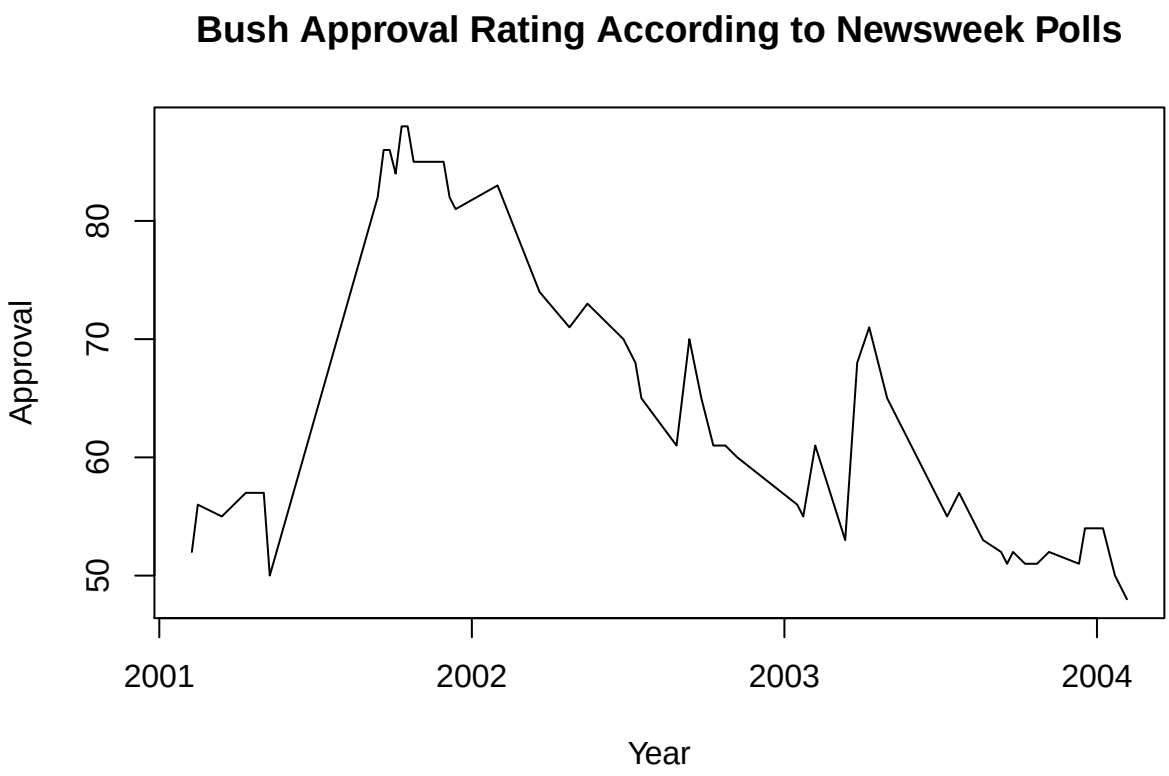
Problem 1

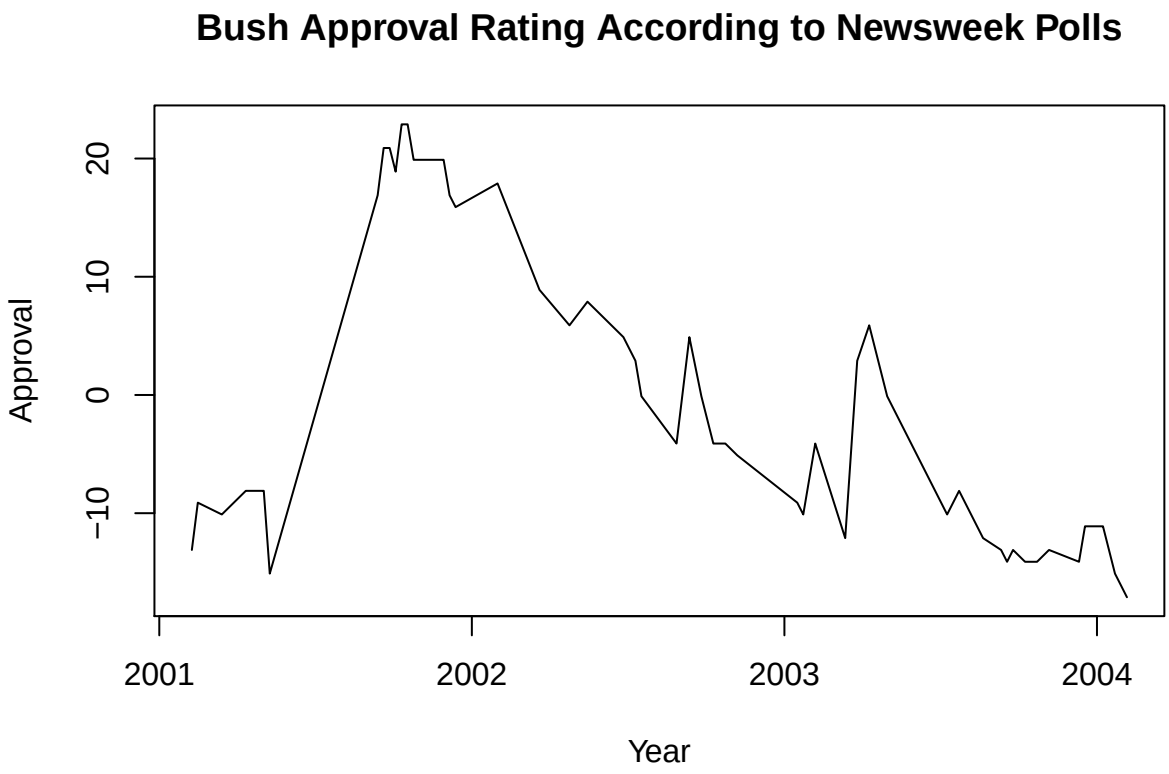
Use the pipe operators from ***magrittr*** *to do the following, all in one pipeline:*

1. *On the data set* `BushApproval` *(**UsingR**), select only Newsweek polls, and work with a data frame that only contains the variables* `date` *and* `approval` *(i.e., drop the variable* `who`*, since it's no longer relevant).*
2. *Make a line plot of Bush's approval rating.* (Hint: The dates are character data, not numeric; consider using `as.Date()`, and set the `format` argument to `"%m/%d/%y"`.)
3. *Subtract out the mean approval rating from the* `approval` *variable, and make a new variable containing the centered ratings.*
4. *Plot the new centered approval ratings as a line plot, as you did in step 2.*

```
# Your code here
suppressMessages(library(UsingR))
library(magrittr)

BushApproval %>%
  subset(select = -who, who == "newsweek") %>%
  transform(date = as.Date(date, format = "%m/%d/%y")) %>%
plot(xlab = "Year", ylab = "Approval", main = "Bush Approval Rating According to Newsweek Polls", type =
  subset(BushApproval, select = -who, who == "newsweek") %>%
  transform(date = as.Date(date, format = "%m/%d/%y")) %>%
  transform(centeredData = approval - mean(approval)) %>%
  subset(select = -approval) %>%
  plot(xlab = "Year", ylab = "Approval", main = "Bush Approval Rating According to Newsweek Polls", type
```


Bush Approval Rating According to Newsweek Polls
Approval
50
60
70
80
2001
2002
2003
2004
Year




```
# bushSubset <- subset(BushApproval, select = -c(who),
#                      subset = who == "newsweek")
#
# bushSubset$date = as.Date(bushSubset$date, "%m/%d/%y")
# plot(bushSubset$date, bushSubset$approval, xlab = "Year", ylab = "Approval", main = "Bush Approval Ra
```

Problem 2

*Use **dplyr** to subset the `Cars93` (**MASS**) data set, selecting only compact cars from the USA, and containing only variables directly related to price and miles per gallon (MPG). Randomly select five observations from the resulting data set.*

```
# Your code here
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:Hmisc':
##
##     src, summarize

## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

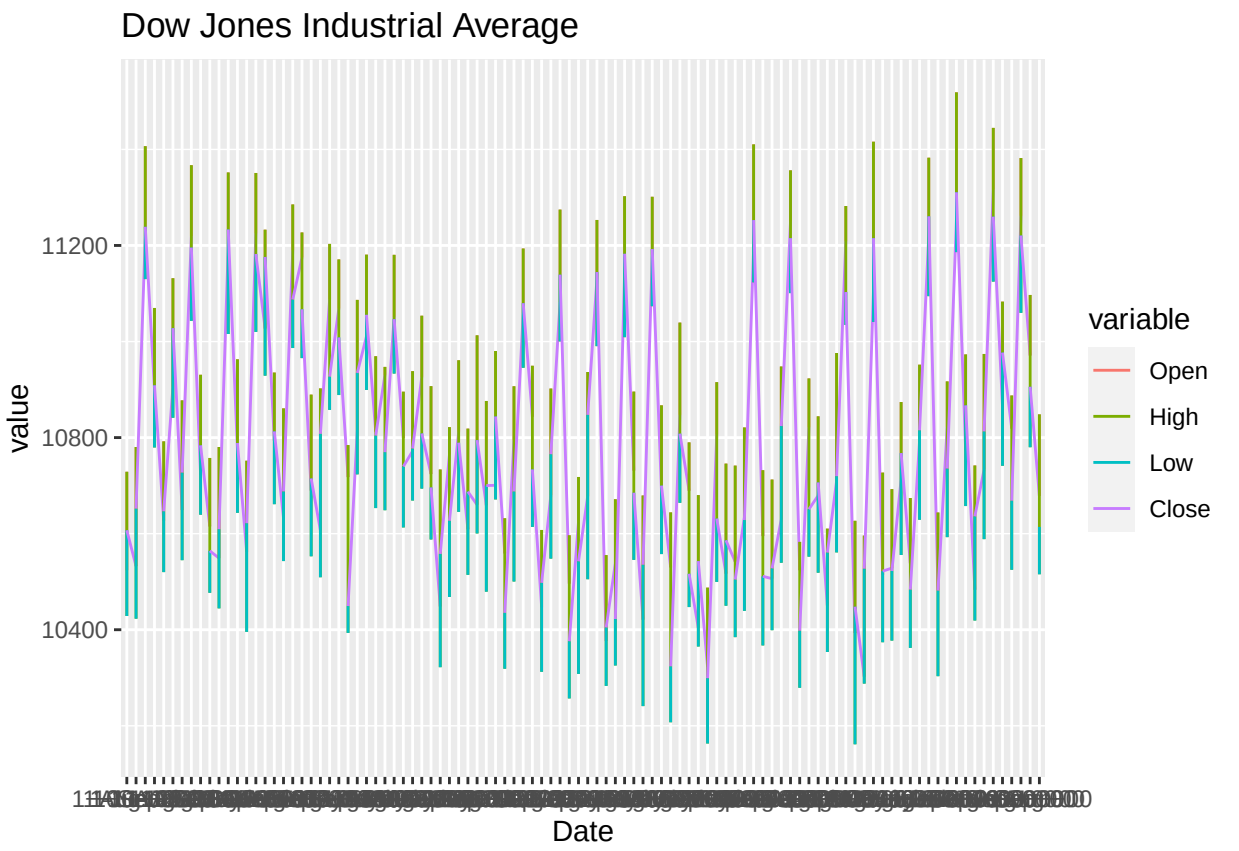
```r
suppressMessages(library(UsingR))
Cars93 %>% filter(Origin == "USA" & Type == "Compact") %>%
  select(Min.Price, Price, Max.Price, MPG.city, MPG.highway) %>% sample(size=5)
```

```
##   Max.Price Min.Price Price MPG.highway MPG.city
## 1      18.3       8.5  13.4          36       25
## 2      11.4      11.4  11.4          34       25
## 3      17.1      14.5  15.8          28       23
## 4      14.7      11.9  13.3          27       22
## 5      12.2      10.4  11.3          27       22
## 6      14.0      13.0  13.5          31       24
## 7      12.8       9.4  11.1          31       23
```

Problem 3

*Use **reshape2** to transform the first 100 rows in the data set `dowdata` (**UsingR**) from wide-form format to long-form format, then plot the open, high, low, and close of the Dow Jones Industrial Average using `ggplot()` from **ggplot2**.*

```r
# Your code here
library(reshape2)
dow = dowdata[1:100,] %>% melt(id.vars = "Date")
ggplot(dow, aes(Date, value, color = variable, group = 1)) + geom_line() +
  labs(x = "Date", y = "value", title = "Dow Jones Industrial Average")
```



Problem 4

Use ***reshape2*** *to transform the data set* `iris` *from wide-form format to long-form format (using* `melt()`*) using* `Species` *as the id variable, then back to wide-form format (using* `dcast()`*), using the mean for aggregation.*

```
# Your code here
library(reshape2)
longDataIris = melt(iris, id.vars = "Species")
wideDataIris = dcast(longDataIris, Species ~ variable, mean)
print(wideDataIris)
```

```
##      Species Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1     setosa        5.006       3.428        1.462       0.246
## 2 versicolor        5.936       2.770        4.260       1.326
## 3  virginica        6.588       2.974        5.552       2.026
```