

Appendix A

Load table in and add a column that calculate the percentage of total followers each subject gained over the 90-day period.

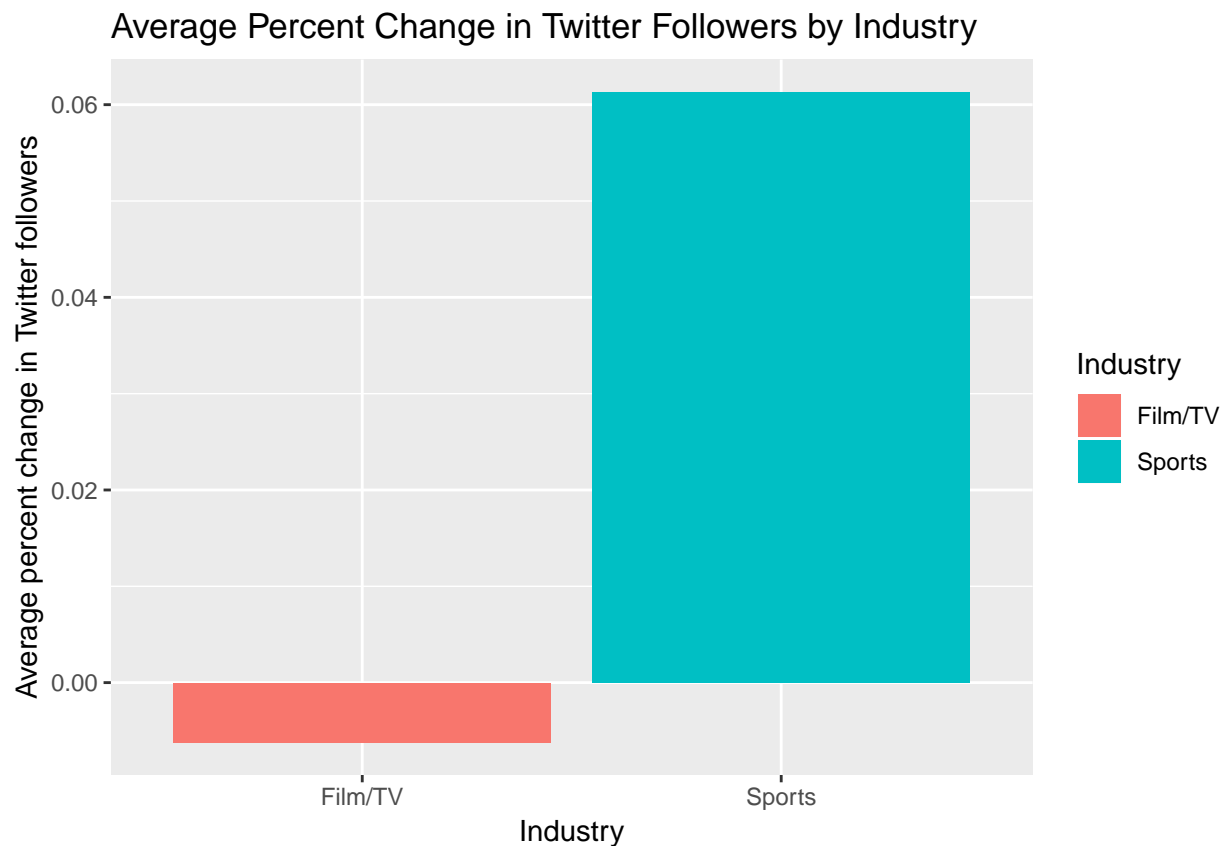
```
tab <- read.csv("TwitterData.csv")

tab <- tab %>%
  mutate(Percent.Follower.Change = 100 * Follower.count.slope.over.90.days.after.scandal / Total.Followers)
```

Group the data by industry, calculate mean percentage change and plot it.

```
new_tab <- tab %>%
  select(Person, Industry, Percent.Follower.Change) %>%
  filter(Person != "Roseanne Barr" & Person != "Daryl Morey") %>%
  group_by(Industry) %>%
  summarise(avg_change = mean(Percent.Follower.Change))

new_tab %>%
  ggplot(aes(x = Industry, y = avg_change, fill=Industry)) +
  geom_bar(stat="identity") +
  ylab("Average percent change in Twitter followers") +
  ggtitle("Average Percent Change in Twitter Followers by Industry")
```



Look at those subjects involved in sexual assault...

```
new_tab <- tab %>%
  filter(Description=="Sexual assault") %>%
  select(Person, Description, Industry, Gain.Lost, Percent.Follower.Change)
```

new_tab

```
##           Person      Description Industry Gain.Lost
## 1      Kevin Spacey Sexual assault  Film/TV      Lost
## 2      Paul Haggis Sexual assault  Film/TV      Lost
## 3 Harvey Weinstein Sexual assault  Film/TV      Lost
## 4      Antonio Brown Sexual assault Sports        Gain
## 5 Cristiano Ronaldo Sexual assault Sports        Gain
## Percent.Follower.Change
## 1          -0.007545891
## 2          -0.003835346
## 3          -0.005377335
## 4           0.111906096
## 5           0.002361641
```

... versus everyone else.

```
new_tab <- tab %>%
  filter(Description!="Sexual assault") %>%
  select(Person, Description, Industry, Gain.Lost, Percent.Follower.Change)
```

new_tab

```
##           Person      Description Industry Gain.Lost
## 1 Jussie Smollett Fake police report  Film/TV      Lost
## 2  Roseanne Barr      Racist tweet  Film/TV      Gain
## 3  Myles Garrett    Physical assault Sports        Gain
## 4 Connor McGregor Destruction of property Sports        Gain
## 5   Daryl Morey      Offensive tweet Sports        Gain
## Percent.Follower.Change
## 1          -0.008094302
## 2           0.287868221
## 3           0.119468196
## 4           0.011484654
## 5           0.295941844
```

Inspect subject's latency against other information, if applicable.

```
new_tab <- tab %>%
  select(Person, Description, Gain.Lost, Latency, Total.Followers) %>%
  filter(!is.na(Latency))
```

new_tab

```
##           Person      Description Gain.Lost Latency Total.Followers
## 1      Kevin Spacey Sexual assault      Lost 17 days      4417411
## 2 Jussie Smollett Fake police report      Lost 42 days      1371335
## 3      Paul Haggis Sexual assault      Lost 25 days       20285
## 4 Harvey Weinstein Sexual assault      Lost 34 days      14464
```

Train a decision tree to predict what percentage change a subject will see, based on the type of scandal (Sexual assault, Tweet-related, or Other) and the subject's industry.

```

input <- tab %>%
  mutate(Sexual.Assault = (Description=="Sexual assault")) %>%
  mutate(Desc =
    ifelse(Description=="Sexual assault",
           "Sexual assault",
           ifelse(Description=="Offensive tweet" | Description=="Racist tweet",
                  "Tweet",
                  "Other")))

binary.model <- rpart(
  Percent.Follower.Change ~ Desc + Industry,
  data = input,
  method = "anova",
  control = rpart.control(minbucket=1, minsplit=1)
)
binary.model

## n= 10
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 10 1.328787e-01  0.080417780
##    2) Desc=Other,Sexual assault 8 2.102894e-02  0.027545960
##      4) Industry=Film/TV 4 1.166747e-05 -0.006213219 *
##      5) Industry=Sports 4 1.189981e-02  0.061305150 *
##    3) Desc=Tweet 2 3.259169e-05  0.291905000 *

rpart.plot(binary.model, extra = 1, type = 3)

```

