Project5

Audrey Holloman and Nicole Kadosh

10/30/2018

2017

```
library(tidyverse)
WP <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/WP.csv")
TM <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/TM.csv")
RO <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/RO.csv")
PO <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/PO.csv")
TDef <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/TDef.csv")
first <- merge(WP, TM, by = "Name")
second <- merge(RO, PO, by = "Name")
ThoseTwo <- merge(first, second, by = "Name")
data <- merge(ThoseTwo, TDef, by = "Name")
# View(data)
attach(data)
data2017 <- data.frame(data$Name, data$Pct, data$Turn.Gain, data$Rush.Yds, data$Pass.Yds, data$YDS)
# View(data2017)
```

```
library(dplyr)
attach(data2017)

data2017$Rank <- (5*Pct) +
  (1*(Turn.Gain/max(Turn.Gain))) +
  (4*(Rush.Yds/max(Rush.Yds))) +
  (2*(Pass.Yds/max(Pass.Yds))) +
  (3*(1 - (YDS/max(YDS))))


#View(data2017)   # the higher the rank, the better the team

# This ranking measure gives a 60% accuracy (from Excel)
```

2016

```
WP2 <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/WP2.csv")
TM2 <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/TM2.csv")
RO2 <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/RO2.csv")
PO2 <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/PO2.csv")
TDef2 <- read.csv("/home/students/kadoshn/GitHub/Sports Analytics/TDef2.csv")

first2 <- merge(WP2, TM2, by = "Name")
second2 <- merge(RO2, PO2, by = "Name")
ThoseTwo2 <- merge(first2, second2, by = "Name")
data2 <- merge(ThoseTwo2, TDef2, by = "Name")
# View(data2)
attach(data2)
data2016 <- data.frame(data2$Name, data2$Pct, data2$Turn.Gain, data2$Rush.Yds, data2$Pass.Yds, data2$YDS
# View(data2016)
```

```
data2016$Rank <- (5*Pct) +
  (1*(Turn.Gain/max(Turn.Gain))) +
  (4*(Rush.Yds/max(Rush.Yds))) +
  (2*(Pass.Yds/max(Pass.Yds))) +
  (3*(1 - (YDS/max(YDS))))


#View(data2016) # the higher the rank the better the team

# This ranking measure gives a 65.85% accuracy (from Excel)
```

After using the same ranking measure equation for both the 2016 and 2017 college football team statistics, we found that our ranking measure for 2016 was 65.85% accurate, and 2017 was 60% accurate. As stated in the project description, the best any polls have done in regard to predicting bowl game results is around 61%. This means that for the year 2016 our ranking measure was more accurate than the polls. On top of that, the year 2017 accuracy for our ranks were about the same as the polls. So overall, our way of ranking the teams in order to predict bowl game outcomes is somewhat better. Granted, we only looked at the last two years individually versus a 10-year window like they did. Also, our ranking algorithm only looks at 5 different metrics.

We chose 5 metrics that we thought would be the most important: winning percentage (Pct), turnover margin (Turn.Gain), rushing offense (Rush.Yds), passing offense (Pass.Yds), and total defense (YDS). In order to scale each of these metrics we had to turn them all into percentages. We did this by dividing the metric by its maximum value for turnover margin, rushing offense, and passing offense. For total defense we also divided the metric by its max, but additionally, we subtracted that value from 1 because the total defensive yards are considered a "negative" statistic. Lastly, we left winning percentage the same since it is already in percent form. To balance our ranking algorithm we used weighted averages. We kept things simple and multiplied each metric percentage by a number 1-5 where 5 is the most important and 1 is of the least importance. This meant that the resulting ranks weren't like normal rank values where the best team is ranked 1 and so on. Instead, the teams with the higher rank mean they are the better team. After obtaining these ranks we then compared them to the bowl game winners for years 2016 and 2017. This is how we found the percentage that corresponds to the accuracy of our ranking algorithm for that particular year. This was done by simply comparing the ranks of the 2 teams that played against one another in each bowl game and finding which team had the better rank between each other (from our ranking algorithm). If the team that won had a better rank from our ranking algorithm, then we wrote a 1 next to that bowl game in Excel. If the team that lost had a better rank, then we wrote a 0 next to that bowl game in Excel. After doing this for every bowl game, we added up the 1's and divided by the number of bowl games to get our percentage of accuracy for each year.