

Does analytics work in professional sports?

Walt DeGrange

October 21, 2019

==

```
library(dplyr)
library(tidyr)
library(magrittr)
library(ggplot2)
library(purrr)
library(ggforce)
library(fpc)
library(cluster)
library(mclust)
library(kableExtra)
library(formattable)
library(stargazer)
library(formatR)
# library(laeres)
```

Sports Analytics - Does it make a difference?

Salary Data - <https://www.spotrac.com>

```
TeamData <- read.csv("ESPNSportsAnalytics.csv")
```

```
colnames(TeamData)[1] <- "Sport"
```

```
TeamData %<>% mutate(PopulationMil = digits(Market/1e+06, 1))
```

```
TeamDataKM2015 <- TeamData %>% select(ESPNRating, X2014HF, X2015HF,
  X2015Salary, PopulationMil)
```

```
TeamDataKM2016 <- TeamData %>% select(ESPNRating, X2014HF, X2015HF,
  X2016Salary, X2016HF, PopulationMil)
```

```
TeamDataKM2017 <- TeamData %>% select(ESPNRating, X2014HF, X2015HF,
  X2016HF, X2017Salary, X2017HF, PopulationMil)
```

```
TeamDataKM2018 <- TeamData %>% select(ESPNRating, X2014HF, X2015HF,
  X2016HF, X2017HF, X2018Salary, X2018HF, PopulationMil)
```

Overall

```
TeamDataNew <- TeamData %>% group_by(Team) %>% mutate(HowFar = sum(X2014HF,
  X2015HF, X2016HF, X2017HF, X2018HF)/5)

TeamDataNew$HowFar[TeamDataNew$Team == "Golden Knights"] <- 5/2

TeamDataNew %>% group_by(Sport) %>% summarise(count = n(), MeanESPN = digits(mean(ESPNRating),
  1), StdESPN = digits(sd(ESPNRating), 1)) %>% kable() %>%
  kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	count	MeanESPN	StdESPN
MLB	30	3.5	1.3
NBA	30	3.1	1.2
NFL	32	2.7	1.0
NHL	31	3.4	0.9

Local Teams

```
TeamDataNew %>% group_by(City, Sport, Team, HowFar) %>% filter(City %in%
  "Raleigh") %>% summarise(count = n(), ESPN = ESPNRating) %>%
  select(City, Team, ESPN, HowFar) %>% kable() %>% kable_styling(c("striped",
  "condensed"), full_width = F) %>% column_spec(column = 1:2,
  color = "black")
```

Sport	City	Team	ESPN	HowFar
NHL	Raleigh	Hurricanes	3	0.6

Champions

2014

```
TeamDataNew %>% filter(X2014HF == "5") %>% select(Sport, Team,
  City, ESPNRating, PopulationMil) %>% arrange(Sport) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil
MLB	Giants	San Francisco	3	0.9
NBA	GSWarriors	Oakland	4	0.4
NFL	Patriots	Boston	4	0.6
NHL	Blackhawks	Chicago	5	2.7

2015

```
TeamDataNew %>% filter(X2015HF == "5") %>% select(Sport, Team,
  City, ESPNRating, PopulationMil) %>% arrange(Sport) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil
MLB	Royals	Kansas City	4	2.2
NBA	Cavaliers	Cleveland	4	0.4
NFL	Broncos	Denver	2	0.6
NHL	Penguins	Pittsburgh	4	0.3

2016

```
TeamDataNew %>% filter(X2016HF == "5") %>% select(Sport, Team,
  City, ESPNRating, PopulationMil) %>% arrange(Sport) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil
MLB	Cubs	Chicago	5	2.7
NBA	GSWarriors	Oakland	4	0.4
NFL	Patriots	Boston	4	0.6
NHL	Penguins	Pittsburgh	4	0.3

2017

```
TeamDataNew %>% filter(X2017HF == "5") %>% select(Sport, Team,
  City, ESPNRating, PopulationMil) %>% arrange(Sport) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil
MLB	Astros	Houston	5	2.1
NBA	GSWarriors	Oakland	4	0.4
NFL	Eagles	Philadelphia	4	1.6
NHL	Capitals	Washington	4	6.1

2018

```
TeamDataNew %>% filter(X2018HF == "5") %>% select(Sport, Team,
  City, ESPNRating, PopulationMil) %>% arrange(Sport) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil
MLB	RedSox	Boston	5	0.6
NBA	Raptors	Toronto	3	2.7
NFL	Patriots	Boston	4	0.6
NHL	Blues	St Louis	4	0.3

Team list grouped by ESPN Ranking

```
TeamDataNew$PopulationMil <- digits(TeamDataNew$PopulationMil,
  1)

TeamDataNew %>% group_by(ESPNRating) %>% summarise(Count = n(),
  HowFar = digits(sum(X2014HF, X2015HF, X2016HF, X2017HF, X2018HF)/(5 *
    n()), 1), MeanPop = digits(mean(PopulationMil), 1), MeanSalary = digits(sum(X2014Salary,
    X2015Salary, X2016Salary, X2017Salary, X2018Salary)/(5 *
    n()), 1)) %>% kable() %>% kable_styling(c("striped",
  "condensed"), full_width = F) %>% column_spec(column = 1:2,
  color = "black")
```

ESPNRating	Count	HowFar	MeanPop	MeanSalary
1	10	0.2	4.3	17.2
2	27	0.7	1.8	17.5
3	34	0.7	0.9	16.8
4	37	1.4	1.7	13.4
5	15	1.2	1.7	14.7

Ten best ESPN Rating performers

```
TeamDataNew %>% select(Sport, Team, City, ESPNRating, PopulationMil,
  HowFar) %>% arrange(-HowFar) %>% head(10, ESPNRating) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil	HowFar
NBA	GSWarriors	Oakland	4	0.4	4.6
NFL	Patriots	Boston	4	0.6	4.4
NBA	Cavaliers	Cleveland	4	0.4	3.4
MLB	Dodgers	Los Angeles	4	3.8	3.0
NHL	Penguins	Pittsburgh	4	0.3	2.8
NBA	Raptors	Toronto	3	2.7	2.6
NHL	Golden Knights	Las Vegas	5	0.6	2.5
NHL	Capitals	Washington	4	6.1	2.4
MLB	Cubs	Chicago	5	2.7	2.2
NHL	Blues	St Louis	4	0.3	2.2

Ten bottom ESPN Rating performers

```
TeamDataNew %>% select(Sport, Team, City, ESPNRating, PopulationMil,
  HowFar) %>% arrange(-ESPNRating, HowFar) %>% head(10) %>%
  kable() %>% kable_styling(c("striped", "condensed"), full_width = F) %>%
  column_spec(column = 1:2, color = "black")
```

Sport	Team	City	ESPNRating	PopulationMil	HowFar
MLB	Rays	Tampa Bay	5	0.3	0.0
NBA	Mavericks	Dallas	5	1.2	0.4
MLB	Pirates	Pittsburgh	5	0.3	0.4
MLB	As	Oakland	5	0.4	0.4
NBA	76ers	Philadelphia	5	1.6	0.8
MLB	Yankees	New York	5	8.6	1.2
NHL	Blackhawks	Chicago	5	2.7	1.4
NBA	Spurs	San Antonio	5	1.5	1.6
MLB	Indians	Cleveland	5	0.4	1.6
MLB	RedSox	Boston	5	0.6	1.8

Data Summary

```
TeamDataDF <- TeamData %>% select(c(ESPNRating,PopulationMil)) %>% data.frame()
ESPNTTable <- stargazer(TeamDataDF, type="html", align = TRUE, digits=1)
```

```
##
## <table style="text-align:center"><tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left">
## <tr><td style="text-align:left">PopulationMil</td><td>123</td><td>1.7</td><td>2.2</td><td>0.1</td><td>0.1</td><td>0.1</td><td>0.1</td></tr><tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr></table>
```

Big to Small Market Teams

Los Angeles

```
TeamDataNew %>% group_by(Sport, Team, HowFar) %>% filter(City %in%
  c("Los Angeles", "Anaheim")) %>% summarise(count = n(), ESPN = (ESPNRating)) %>%
  select(Team, ESPN, HowFar) %>% kable() %>% kable_styling(c("striped",
  "condensed"), full_width = F) %>% column_spec(column = 1:2,
  color = "black")
```

Sport	Team	ESPN	HowFar
MLB	Angels	3	0.4
MLB	Dodgers	4	3.0
NBA	Clippers	2	1.0
NBA	Lakers	1	0.0
NFL	Chargers	1	0.4
NFL	Rams	2	1.0
NHL	Ducks	2	1.6
NHL	Kings	4	0.4

New York

```
TeamDataNew %>% group_by(Sport, Team, HowFar) %>% filter(City %in%
  c("New York")) %>% summarise(count = n(), ESPN = (ESPNRating)) %>%
  select(Team, ESPN, HowFar) %>% kable() %>% kable_styling(c("striped",
  "condensed"), full_width = F) %>% column_spec(column = 1:2,
  color = "black")
```

Sport	Team	ESPN	HowFar
MLB	Mets	4	1.0
MLB	Yankees	5	1.2
NBA	Knicks	1	0.2
NBA	Nets	1	0.2
NFL	Giants	2	1.6
NFL	Jets	1	1.0
NHL	Islanders	4	1.0
NHL	Rangers	2	2.0

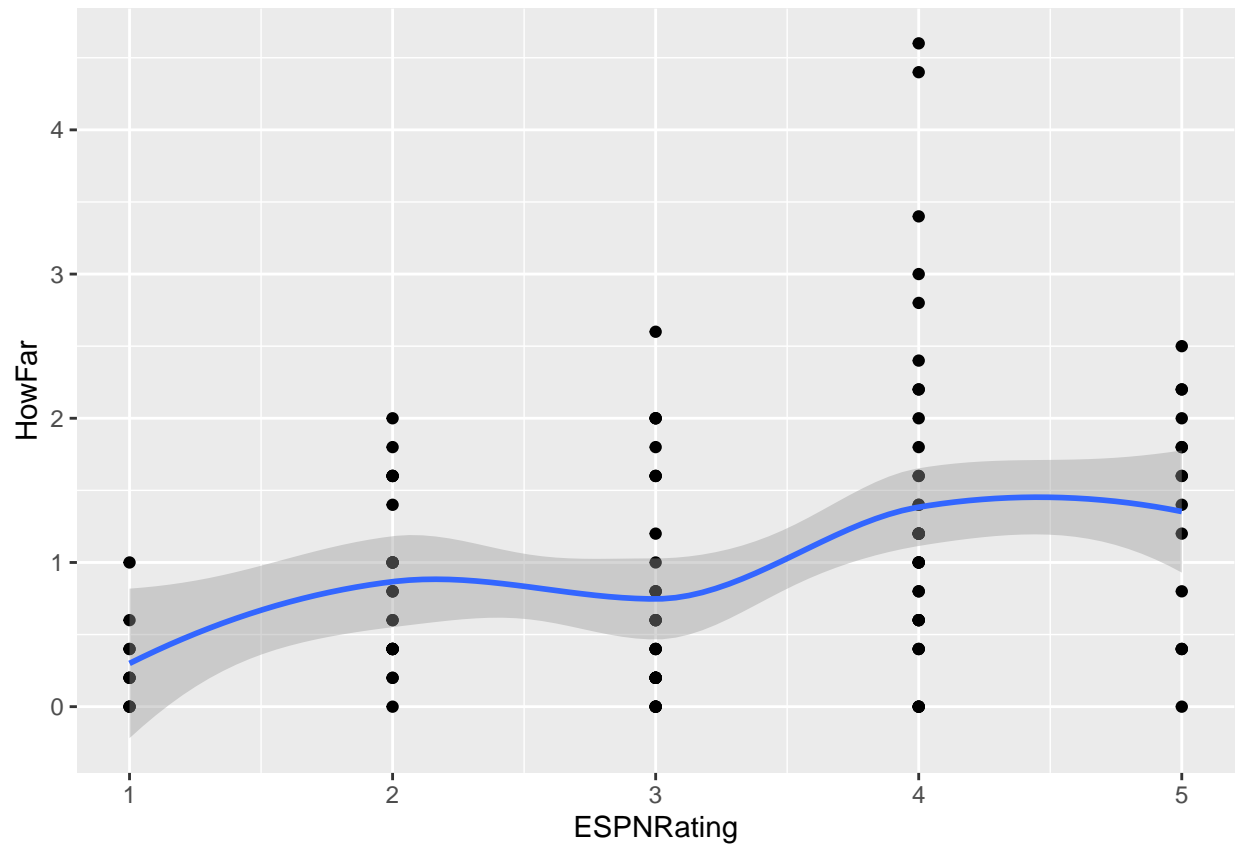
Green Bay

```
TeamDataNew %>% group_by(Sport, Team, HowFar) %>% filter(City %in%
  c("Green Bay")) %>% summarise(count = n(), ESPN = (ESPNRating)) %>%
  select(Team, ESPN, HowFar) %>% kable() %>% kable_styling(c("striped",
    "condensed"), full_width = F) %>% column_spec(column = 1:2,
    color = "black")
```

Sport	Team	ESPN	HowFar
NFL	Packers	3	1.6

Plot of ESPN vs How Far a team made it into the playoffs on average for the 4 years of data

```
TeamDataNew %>% ggplot(aes(x = ESPNRating, y = HowFar)) + geom_point() +
  geom_smooth()
```



K Means clustering 2018

```
TeamDataKM2018 <- scale(TeamDataKM2018)

fit <- kmeans(TeamDataKM2018, 2)

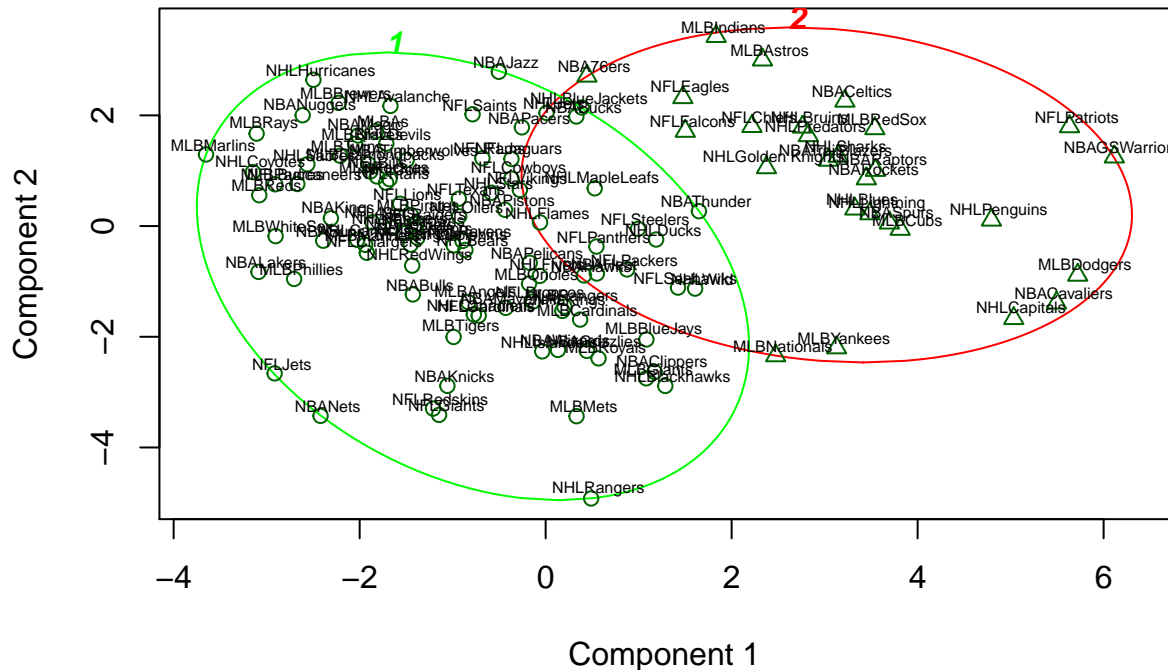
TeamDataKM2018 <- data.frame(TeamDataKM2018, fit$cluster)

TeamData2018 <- data.frame(TeamData, fit$cluster)

rownames(TeamData2018) <- paste(TeamData2018$Sport, TeamData2018$Team,
  sep = "")

clusplot(TeamData2018, fit$cluster, color = TRUE, shade = FALSE,
  labels = 2, lines = 0, cex.txt = 0.5, col.clus = c("red",
    "green"), col.txt = c("black", "black"))
```

CLUSPLOT(TeamData2018)



These two components explain 34.15 % of the point variability.

```
TeamData2018 %>% group_by(TeamData2018$fit.cluster) %>% summarise(count = n(),
  MeanX2014HF = digits(mean(X2014HF), 1), MeanX2015HF = digits(mean(X2015HF),
  1), MeanX2016HF = digits(mean(X2016HF), 1), MeanX2017HF = digits(mean(X2017HF),
  1), MeanX2018HF = digits(mean(X2018HF), 1), MeanPopulation = digits(mean(PopulationMil),
  1), MeanSalary = digits(mean(X2018Salary), 1), MeanESPNRating = digits(mean(ESPNRating),
  1)) %>% kable(format = "latex") %>% kable_styling(c("striped",
  "condensed"), full_width = F, latex_options = "scale_down") %>%
  column_spec(column = 1:2, color = "black")
```

TeamData2018\$fit.cluster	count	MeanX2014HF	MeanX2015HF	MeanX2016HF	MeanX2017HF	MeanX2018HF	MeanPopulation	MeanSalary	MeanESPNRating
1	96	0.8	0.7	0.5	0.4	0.5	1.7	17.4	2.9
2	27	1.3	1.9	2.3	2.7	2.3	1.8	10.4	4.2

K-Means Clustering 2017

```
TeamDataKM2017 <- scale(TeamDataKM2017)

fit <- kmeans(TeamDataKM2017, 2)

TeamDataKM2017 <- data.frame(TeamDataKM2017, fit$cluster)

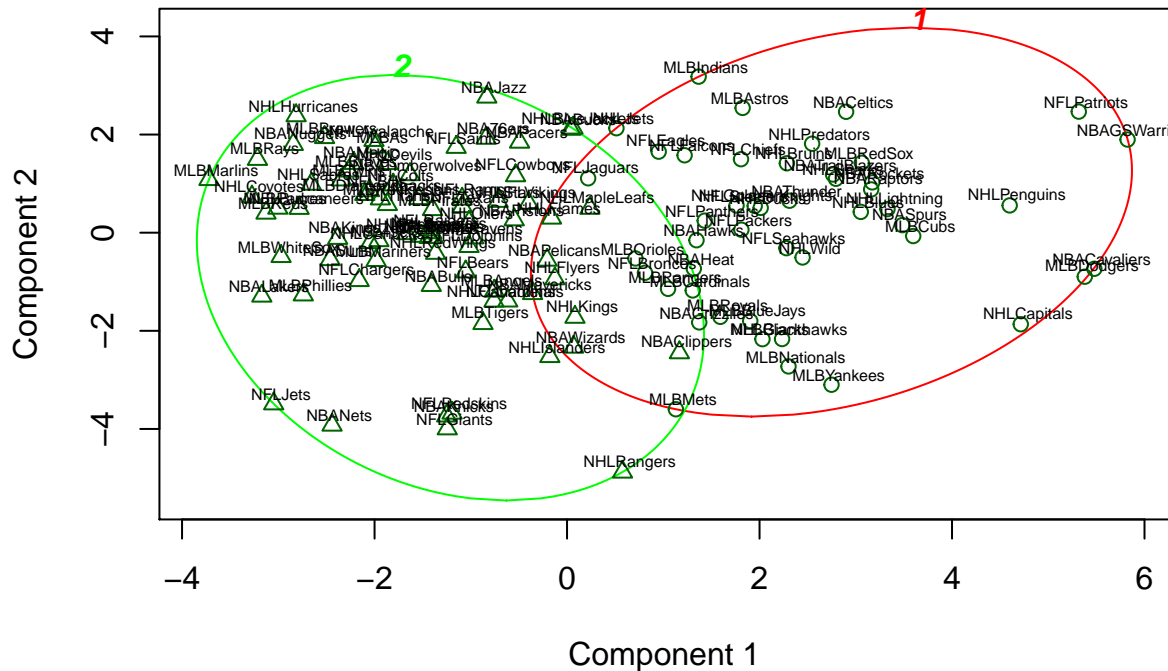
TeamData2017 <- data.frame(TeamData, fit$cluster)

rownames(TeamData2017) <- paste(TeamData2017$Sport, TeamData2017$Team,
  sep = " ")
```



```
clusplot(TeamData2017, fit$cluster, color = TRUE, shade = FALSE,
labels = 2, lines = 0, cex.txt = 0.5, col.clus = c("red",
"green"), col.txt = c("black", "black"))
```

CLUSPLOT(TeamData2017)



These two components explain 34.26 % of the point variability.

```
TeamData2017 %>% group_by(TeamData2017$fit.cluster) %>% summarise(count = n(),
  MeanX2014HF = digits(mean(X2014HF), 1), MeanX2015HF = digits(mean(X2015HF),
  1), MeanX2016HF = digits(mean(X2016HF), 1), MeanX2017HF = digits(mean(X2017HF),
  1), MeanPopulation = digits(mean(PopulationMil), 1),
  MeanSalary = digits(mean(X2017Salary), 1), MeanESPNRating = digits(mean(ESPNRating),
  1)) %>% kable(format = "latex") %>% kable_styling(c("striped",
"condensed"), full_width = F, latex_options = "scale_down") %>%
column_spec(column = 1:2, color = "black")
```

TeamData2017\$fit.cluster	count	MeanX2014HF	MeanX2015HF	MeanX2016HF	MeanX2017HF	MeanPopulation	MeanSalary	MeanESPNRating
1	47	1.6	2.0	1.8	1.8	1.6	10.0	3.9
2	76	0.5	0.3	0.4	0.4	1.8	19.5	2.7

K Means clustering 2016

```
TeamDataKM2016 <- scale(TeamDataKM2016)

fit <- kmeans(TeamDataKM2016, 2)

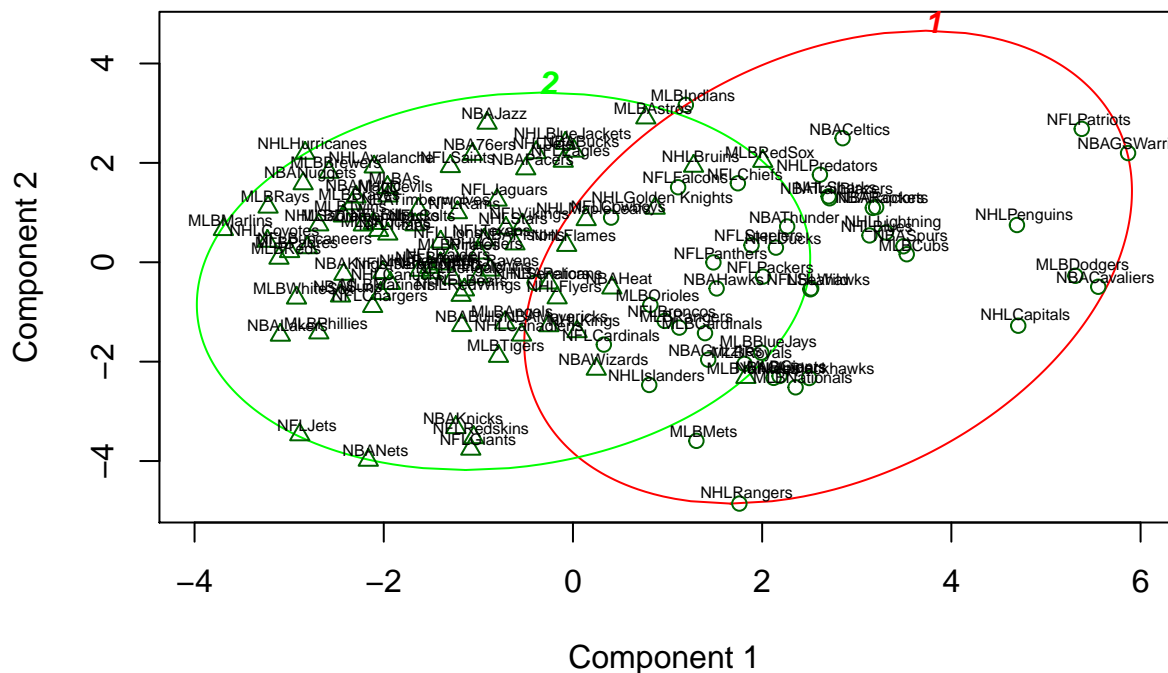
TeamDataKM2016 <- data.frame(TeamDataKM2016, fit$cluster)

TeamData2016 <- data.frame(TeamData, fit$cluster)
```

```
rownames(TeamData2016) <- paste(TeamData2016$Sport, TeamData2016$Team,
  sep = "")

clusplot(TeamData2016, fit$cluster, color = TRUE, shade = FALSE,
  labels = 2, lines = 0, cex.txt = 0.5, col.clus = c("red",
  "green"), col.txt = c("black", "black"))
```

CLUSPLOT(TeamData2016)



These two components explain 33.92 % of the point variability.

```
TeamData2016 %>% group_by(TeamData2016$fit.cluster) %>% summarise(count = n(),
  MeanX2014HF = digits(mean(X2014HF), 1), MeanX2015HF = digits(mean(X2015HF),
  1), MeanX2016HF = digits(mean(X2016HF), 1), MeanPopulation = digits(mean(PopulationMil),
  1), MeanSalary = digits(mean(X2016Salary), 1), MeanESPNRating = digits(mean(ESPNRating),
  1)) %>% kable(format = "latex") %>% kable_styling(c("striped",
  "condensed"), full_width = F, latex_options = "scale_down") %>%
  column_spec(column = 1:2, color = "black")
```

TeamData2016\$fit.cluster	count	MeanX2014HF	MeanX2015HF	MeanX2016HF	MeanPopulation	MeanSalary	MeanESPNRating
1	44	2.0	2.1	2.1	1.9	13.8	3.6
2	79	0.4	0.3	0.3	1.6	16.7	2.9

K Means clustering 2015

```
TeamDataKM2015 <- scale(TeamDataKM2015)

fit <- kmeans(TeamDataKM2015, 2)
```

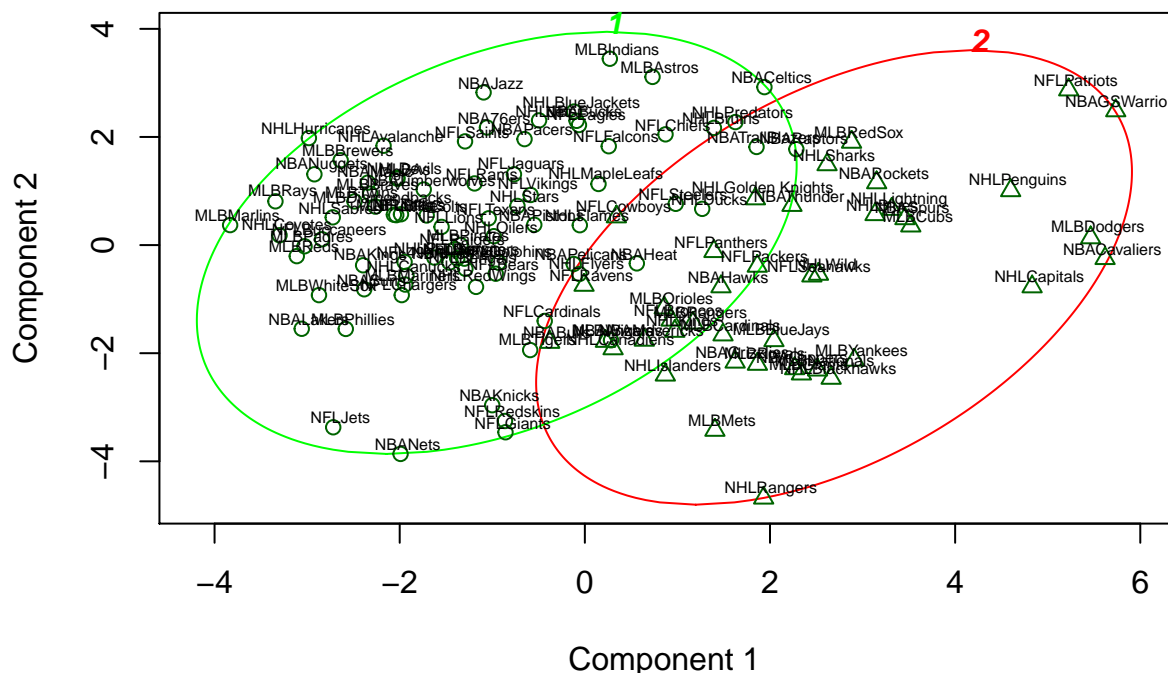
```
TeamDataKM2015 <- data.frame(TeamDataKM2015, fit$cluster)

TeamData2015 <- data.frame(TeamData, fit$cluster)

rownames(TeamData2015) <- paste(TeamData2015$Sport, TeamData2015$Team,
  sep = "")

clusplot(TeamData2015, fit$cluster, color = TRUE, shade = FALSE,
  labels = 2, lines = 0, cex.txt = 0.5, col.clus = c("red",
  "green"), col.txt = c("black", "black"))
```

CLUSPLOT(TeamData2015)



These two components explain 33.91 % of the point variability.

```
TeamData2015 %>% group_by(TeamData2015$fit.cluster) %>% summarise(count = n(),
  MeanX2014HF = digits(mean(X2014HF), 1), MeanX2015HF = digits(mean(X2015HF),
  1), MeanPopulation = digits(mean(PopulationMil), 1),
  MeanSalary = digits(mean(X2015Salary), 1), MeanESPNRating = digits(mean(ESPNRating),
  1)) %>% kable(format = "latex") %>% kable_styling(c("striped",
  "condensed"), full_width = F, latex_options = "scale_down") %>%
  column_spec(column = 1:2, color = "black")
```

TeamData2015\$fit.cluster	count	MeanX2014HF	MeanX2015HF	MeanPopulation	MeanSalary	MeanESPNRating
1	81	0.4	0.4	1.5	19.3	2.8
2	42	2.0	1.9	2.2	8.5	3.8

```
ESPNAnalytics <- data.frame(Year = c(2015, 2016, 2017, 2018),
  High = c(3.4, 3.6, 3.9, 3.9), Low = c(3, 2.9, 2.7, 2.8))
```

```
ESPNAalytics %>% ggplot() + geom_line(aes(x = Year, y = High,
color = "Green")) + geom_line(aes(x = Year, y = Low), color = "Dark Blue") +
scale_x_continuous(breaks = seq(2015, 2018, by = 1)) + scale_y_continuous(limits = c(0,
5)) + ylab("Mean K-Means Group ESPNAalytics Score") + theme(legend.position = "none")
```

