Lab 3

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────────────────────────────────────── tidyv
erse 1.2.1 ──
```

```
## ✔ ggplot2 3.2.1     ✔ purrr   0.3.2
## ✔ tibble  2.1.3     ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0     ✔ stringr 1.4.0
## ✔ readr   1.3.1     ✔ forcats 0.4.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────────────────────── tidyverse_c
onflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
bball = read.table("https://raw.githubusercontent.com/kbodwin/ShinyLabs/master/Datasets/cp_bball.csv", heade
r = TRUE)
```

The Dataset

```
bball %>% tail()
```

```
##         Date     Team Team.Location Team.Score               Opponent
## 22  2/20/16 Cal Poly          Home         71   Cal State Northridge
## 23  2/25/16 Cal Poly          Home         77    Cal State Fullerton
## 24   3/3/16 Cal Poly          Away         62              UC-Irvine
## 25   3/5/16 Cal Poly          Away         50       UC-Santa Barbara
## 26 11/18/15 Cal Poly          Home         71 Cal State Monterey Bay
## 27 11/28/15 Cal Poly          Home         85        Antelope Valley
##    Opponent.Score Team.Result
## 22             75        Loss
## 23             78        Loss
## 24             72        Loss
## 25             69        Loss
## 26             62         Win
## 27             55         Win
```

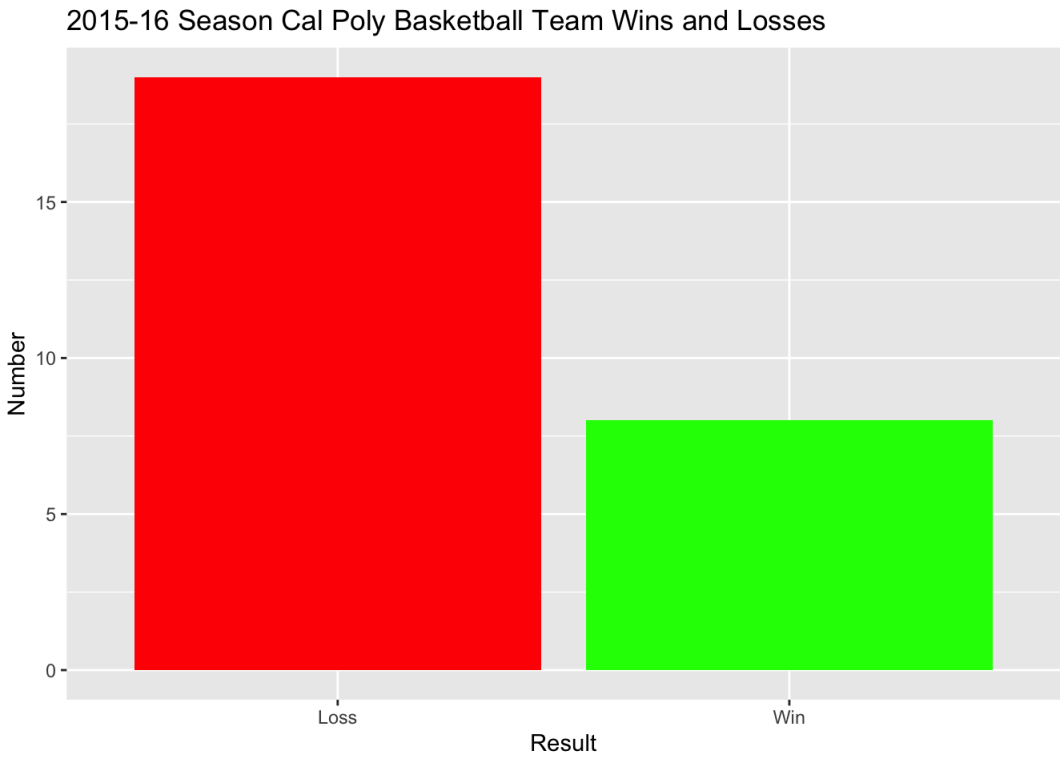
The dataset has information about the Cal Poly basketball team from the 2015-2016 season. This includes info such as their opponents, points scored by both teams, game dates, their locations, results, and more.

Is Cal Poly a worse than average team?

```
bball %>% count(Team.Result)
```

```
## # A tibble: 2 x 2
##   Team.Result     n
##   <fct>       <int>
## 1 Loss           19
## 2 Win             8
```

```
ggplot(bball, aes(x = Team.Result)) +
  geom_bar(fill = c("red","green")) +
  ggtitle("2015-16 Season Cal Poly Basketball Team Wins and Losses") +
  xlab("Result") +
  ylab("Number")
```



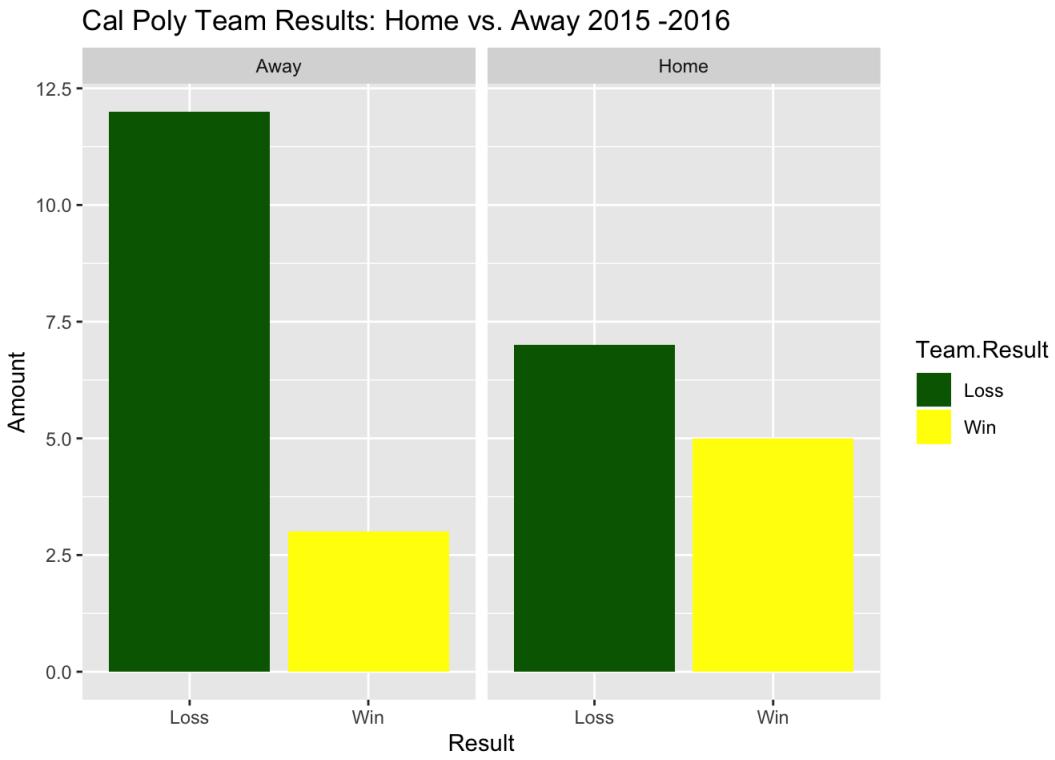

```
a = 8/27
paste("Win Percentage = ", a)
```

```
## [1] "Win Percentage =  0.296296296296296"
```

In the 2015-2016 season, Cal Poly won 8 games and lost 19 games. This means they won 29.63% of their games. Cal Poly is not an average team; they are a below average team since they won below significantly below than 50% of their games.

Was Cal Poly more likely to win at home more than away?

```
ggplot(bball, aes(x = Team.Result, fill = Team.Result)) +
  geom_bar() +
  facet_grid(~Team.Location) +
  ggtitle(label = "Cal Poly Team Results: Home vs. Away 2015 -2016") +
  scale_fill_manual(values = c("dark green", "yellow")) +
  xlab("Result") +
  ylab("Amount")
```




```
bball %>% count(Team.Location, Team.Result)
```

```
## # A tibble: 4 x 3
##   Team.Location Team.Result     n
##   <fct>         <fct>       <int>
## 1 Away          Loss           12
## 2 Away          Win             3
## 3 Home          Loss            7
## 4 Home          Win             5
```

```
3/15
```

```
## [1] 0.2
```

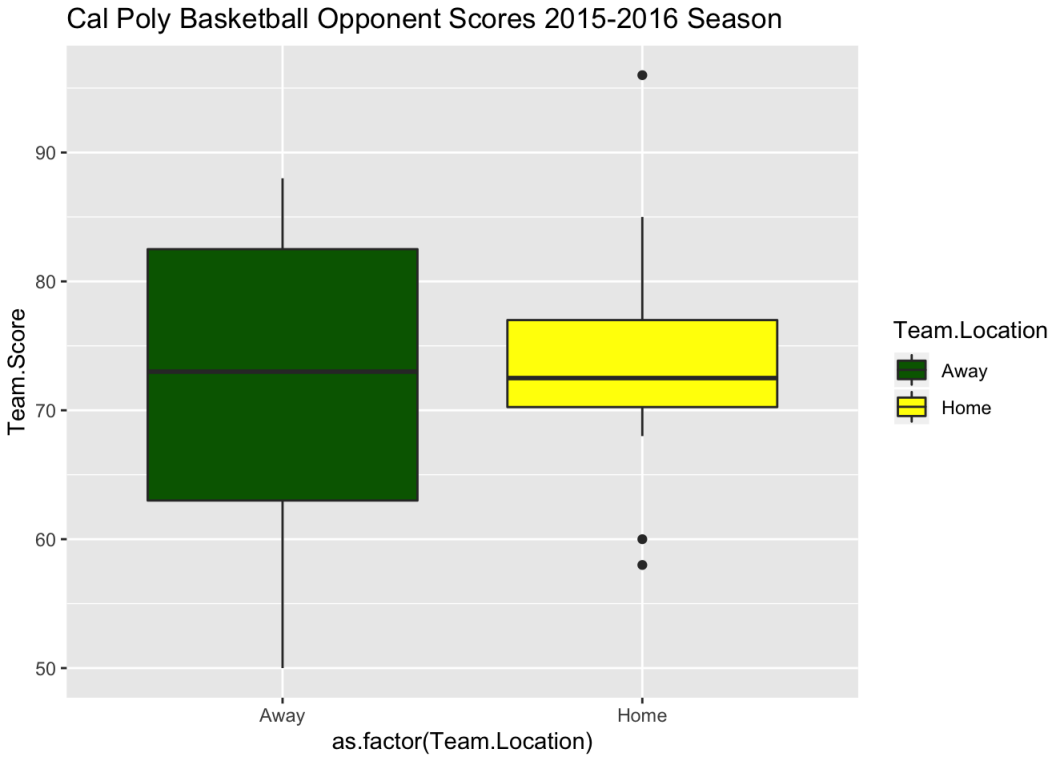
```
5/12
```

```
## [1] 0.4166667
```

Cal Poly won 41.67% of its home games and 20% of its away games. Therefore, Cal Poly is much more likely to win at home than away. Home court advantage and support, along with other external factors, probably has a positive impact on their play at home compared to on the road.

Does Cal Poly score more at home than away?

```
ggplot(bball, aes(x = as.factor(Team.Location), y = Team.Score, fill = Team.Location)) +
  geom_boxplot() +
  scale_fill_manual(values = c("dark green", "yellow")) +
  ggtitle("Cal Poly Basketball Opponent Scores 2015-2016 Season")
```




```
  xlab("Team Location")
```

```
## $x
## [1] "Team Location"
##
## attr(,"class")
## [1] "labels"
```

```
  ylab("Team Score")
```

```
## $y
## [1] "Team Score"
##
## attr(,"class")
## [1] "labels"
```

```
bball %>% group_by(Team.Location) %>% summarize_at(vars(Team.Score), funs(mean))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
## # A tibble: 2 x 2
##   Team.Location Team.Score
##   <fct>              <dbl>
## 1 Away                71.7
## 2 Home                73.4
```

```
bball %>% group_by(Team.Location) %>% summarize_at(vars(Team.Score), funs(sd))
```

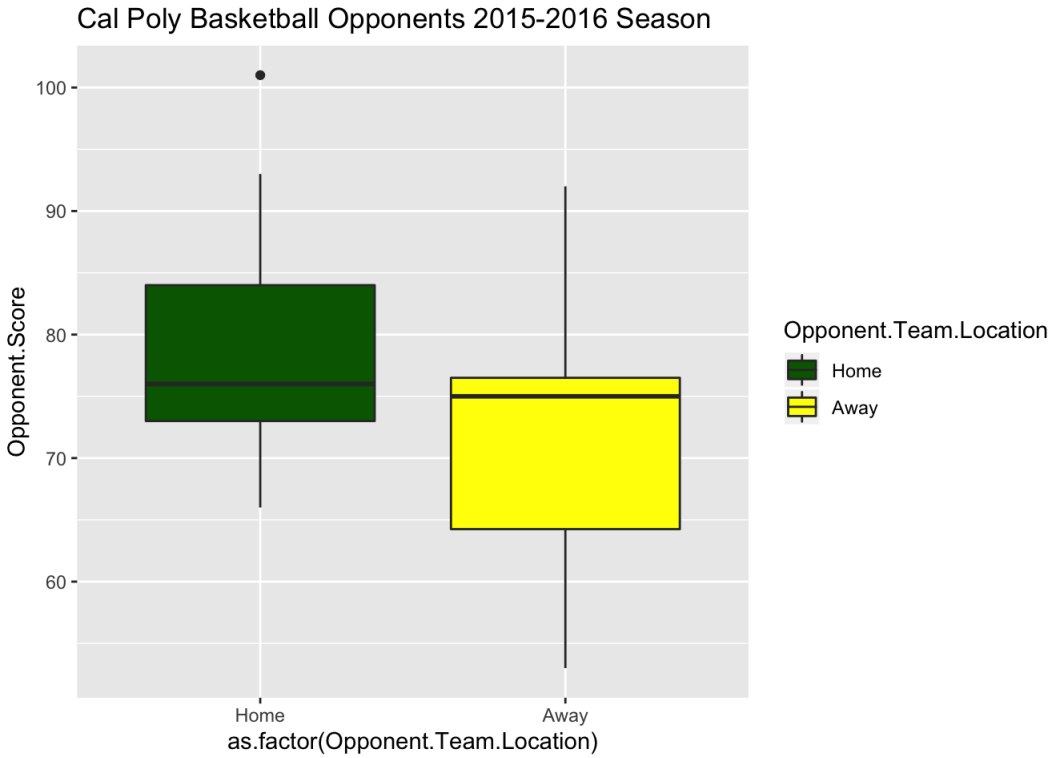
```
## # A tibble: 2 x 2
##   Team.Location Team.Score
##   <fct>              <dbl>
## 1 Away                11.8
## 2 Home                10.1
```

In Home Games, Cal Poly scored an average of 73.42 points with a standard deviation of 10.14 points. In Away Games, Cal Poly scored an average of 71.67 points with a standard deviation of 11.81 points. As a result, Cal Poly is a much better team on offense at Home compared to on the road where they played 3 more games.

Does Cal Poly truly play better at home?

```
bball <- bball %>%
  mutate(
    Opponent.Team.Location = factor(Team.Location,
                                    labels = c("Home", "Away")
    )
    )
```

```
ggplot(bball, aes(x = as.factor(Opponent.Team.Location), y = Opponent.Score, fill = Opponent.Team.Location))
+
  geom_boxplot() +
  scale_fill_manual(values = c("dark green", "yellow")) +
  ggtitle("Cal Poly Basketball Opponents 2015-2016 Season")
```




```
  xlab("Opponent Team Location")
```

```
## $x
## [1] "Opponent Team Location"
##
## attr(,"class")
## [1] "labels"
```

```
  ylab("Score")
```

```
## $y
## [1] "Score"
##
## attr(,"class")
## [1] "labels"
```

“Even if Cal Poly didn’t score much more at home than away, they still played better. This is because they played better defense at home, so the scores overall were lower.” After the analysis of our opponents teams’ scoring, it seems that this statement is accurate. Although Cal Poly doesn’t score may not score much more, their defense seems to improve a lot. Cal Poly holds their opponent to about 71.3 points instead of the 79 points average for away games. The 71.33 is actually lower than their scoring average, which means in total they are scoring more than they are getting scored on. No team scored 100 on them at home, unlike on the road. This would account for why they win about 20% more home games than away games.