

Homework 1 - Data Visualization

Hunter Geise

2-6-2022

Overview and Instructions

This problem set is intended to improve your skills using `ggplot2` and basic data structure manipulation. In each case below, you will see a graphic. Please provide code where you see **# Your code goes here** to recreate the following plots to the best of your ability, with line breaks as appropriate. Echo all your code.

You may work together and seek peer help on this project. However, everyone's solutions must be their unique code. Your answers must be typeset in LaTeX using `knitr` as described in lecture (just use the template). If you have any questions, please contact me.

Download the `ncaaSeason.csv` file from Blackboard. Feel free to examine the data to get a feel for the information contained within the file.

Questions

1. Use `ncaa_18_high` until otherwise noted.

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

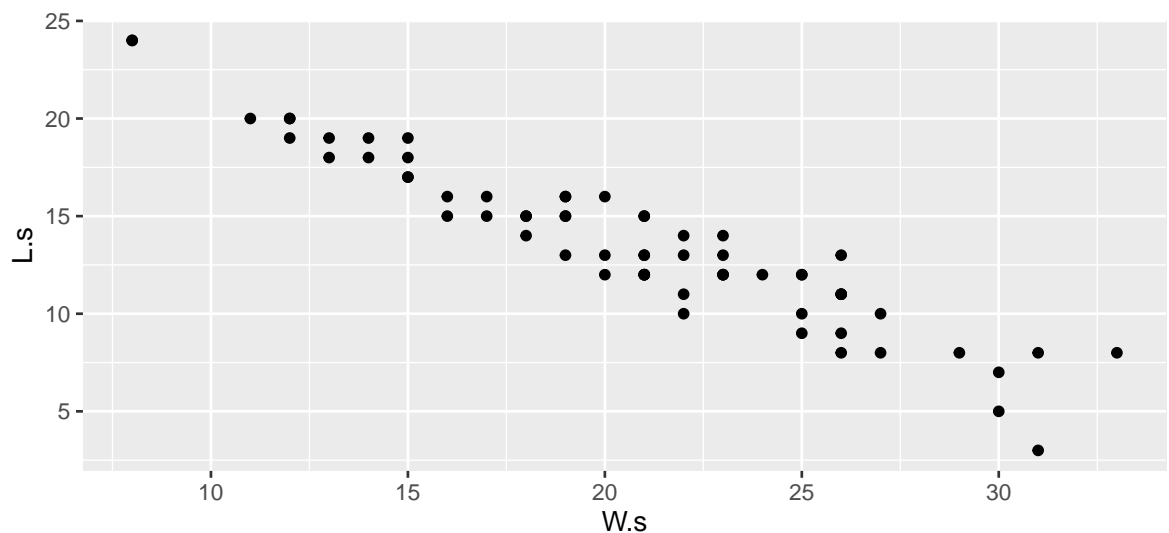
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1

## Warning: package 'tibble' was built under R version 4.1.2

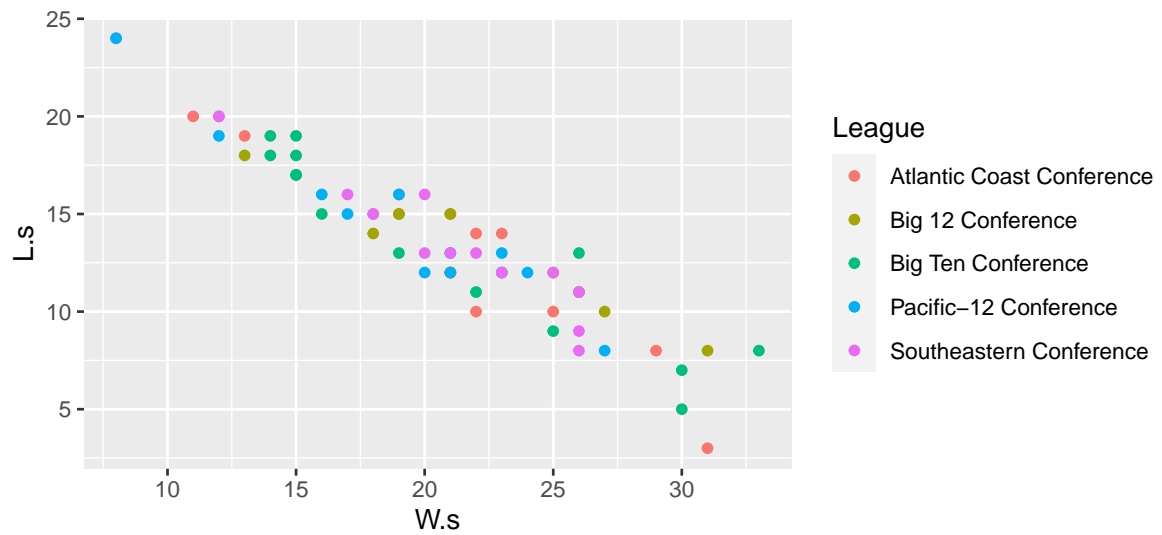
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

# Make sure the data file is located in the same folder as
# your .Rmd file.
ncaa <- read.csv("ncaaSeason.csv")
ncaa_18_high <- ncaa %>%
  filter(Season == 2018,
         League %in% c("Atlantic Coast Conference",
                       "Big 12 Conference",
                       "Big Ten Conference",
                       "Pacific-12 Conference",
                       "Southeastern Conference"))

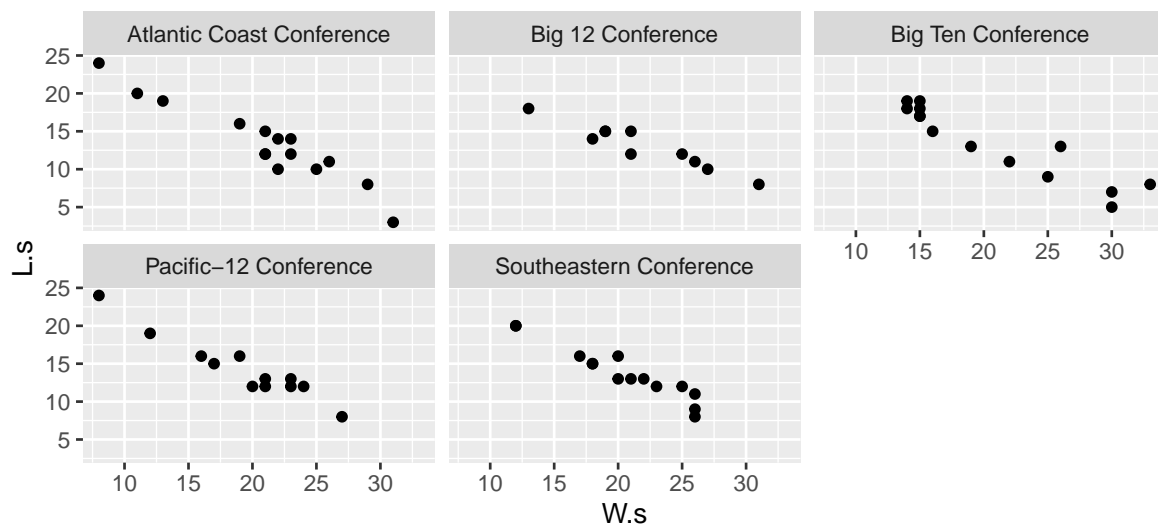
ggplot(data = ncaa_18_high, aes(x = W.s, y = L.s)) +
  geom_point()
```



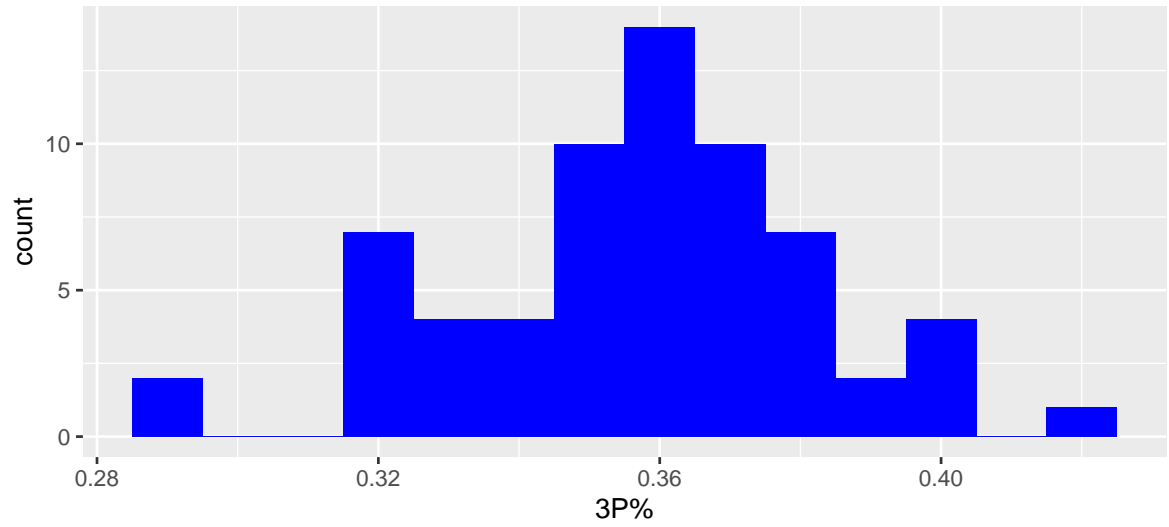
```
2. ggplot(ncaa_18_high, aes(W.s, L.s, color = League)) +  
  geom_point()
```



```
3. ggplot(ncaa_18_high, aes(W.s, L.s)) +  
  geom_point() +  
  facet_wrap(vars(League))
```

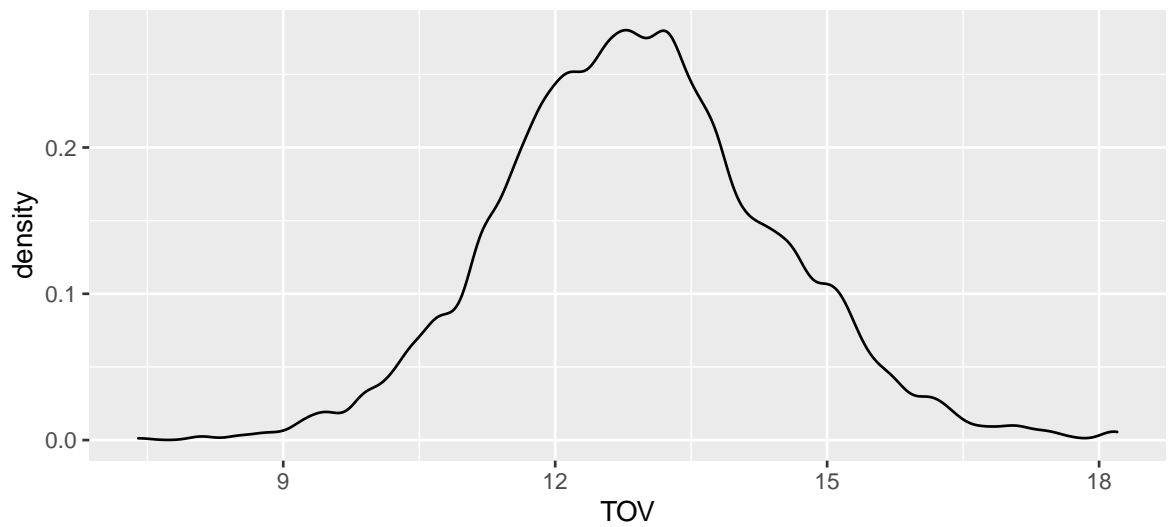


```
4. #check this one
ggplot(ncaa_18_high, aes(X3P.)) +
  geom_histogram(bins = 14, fill = "Blue") +
  labs(x = "3P%")
```

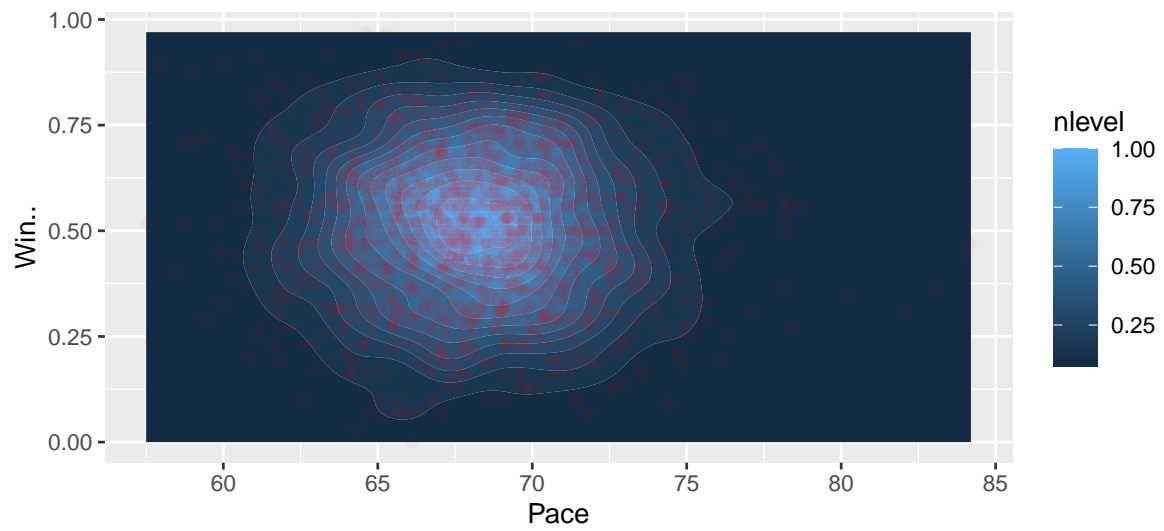


5. Use `ncaa` rather than `ncaa_18_high` until otherwise noted.

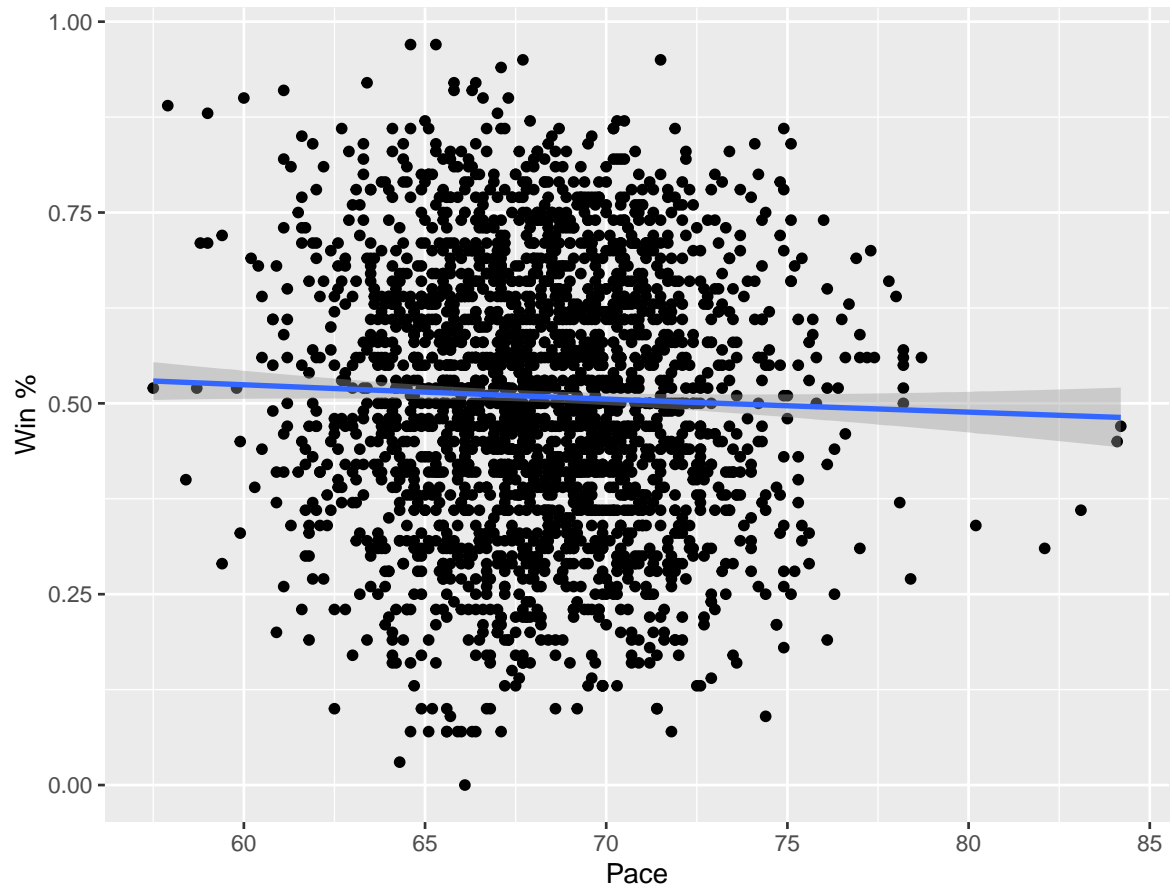
```
#check this one  
ggplot(ncaa, aes(TOV)) +  
  geom_density(adjust = .5)
```



```
6. ggplot(ncaa, aes(Pace, Win..)) +  
  geom_density2d_filled(aes(fill = stat(nlevel))) +  
  geom_point(color = alpha("red", 0.05))
```



```
7. ggplot(ncaa, aes(Pace, Win..)) +  
  geom_point() +  
  geom_smooth() +  
  labs(y = "Win %")
```




```
8. ggplot(ncaa, aes(GP, W.s)) +
  labs(y = "Wins") +
  theme_bw() +
  stat_summary_bin(
    fun.min = min, fun.max = max,
    geom = "errorbar"
  )
```

