

MATH 3070 Lab Project 7

Aini Liang

September 29, 2016

- [Problem 1 \(Verzani problem 5.6\)](#)
- [Problem 2 \(Verzani problem 5.7\)](#)
- [Problem 3](#)
- [BONUS Problem](#)

Remember: I expect to see commentary either in the text, in the code with comments created using `#`, or (preferably) both! **Failing to do so may result in lost points!**

Problem 1 (Verzani problem 5.6)

For the `batting` (**UsingR**) data set, make parallel boxplots of the batting average (`H/AB`) for each team. Which team had the greatest median average? (Use **`lattice`** functions for this problem.)

```
# Your code here
```

```
library(UsingR)
```

```
## Loading required package: MASS
```

```
## Loading required package: HistData
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##  
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':  
##  
##      format.pval, round.POSIXt, trunc.POSIXt, units
```

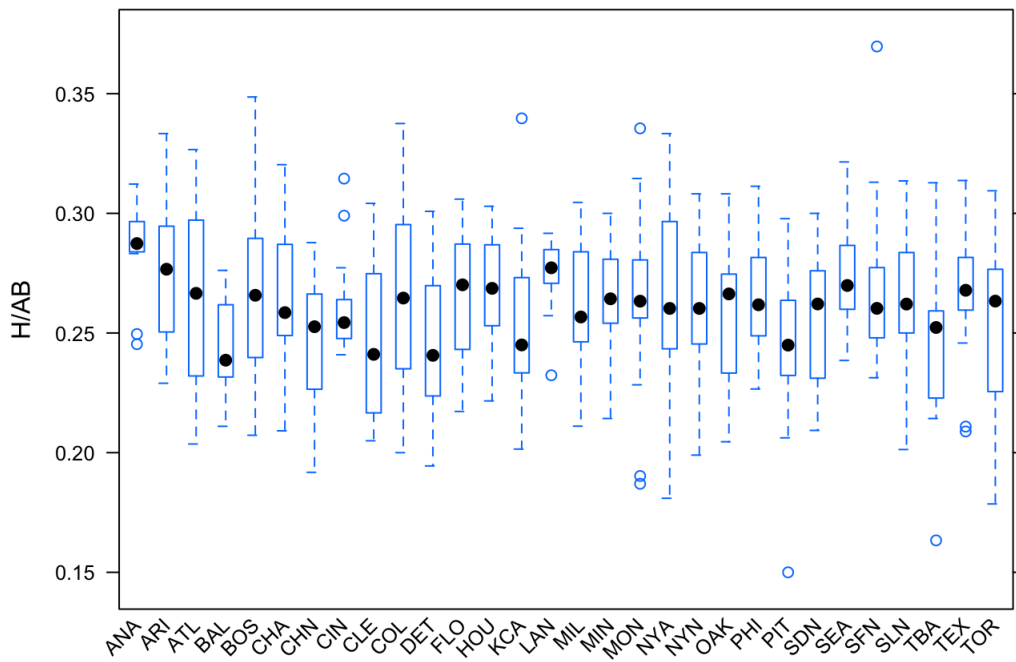
```
##  
## Attaching package: 'UsingR'
```

```
## The following object is masked from 'package:survival':  
##  
##      cancer
```

```
View(batting)
```

```
## Error in check_for_XQuartz(): X11 library is missing: install XQuartz from xquartz.macosforge.org
```

```
bwplot(H/AB ~ teamID, data = batting, scales = list(x = list(rot = 45)))
```



Problem 2 (Verzani problem 5.7)

For the `mtcars` data set, produce graphics of the following using `ggplot2`:

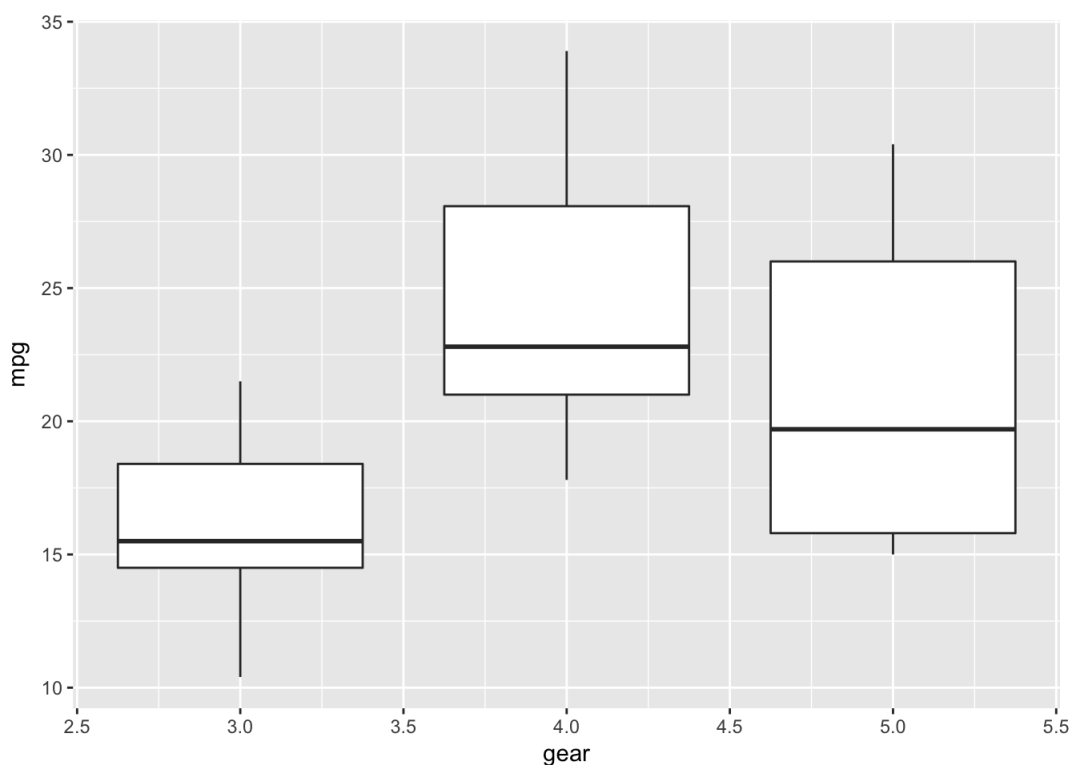
1. Boxplots for miles per gallon (`mpg`) for groups defined by the number of gears (`gear`).

```
# Your code here
```

```
library(ggplot2)
View(mtcars)
```

```
## Error in check_for_XQuartz(): X11 library is missing: install XQuartz from xquartz.macosforge.org
```

```
p <- ggplot(mtcars, aes(x = gear, y = mpg, group = gear)) + geom_boxplot()
print(p)
```

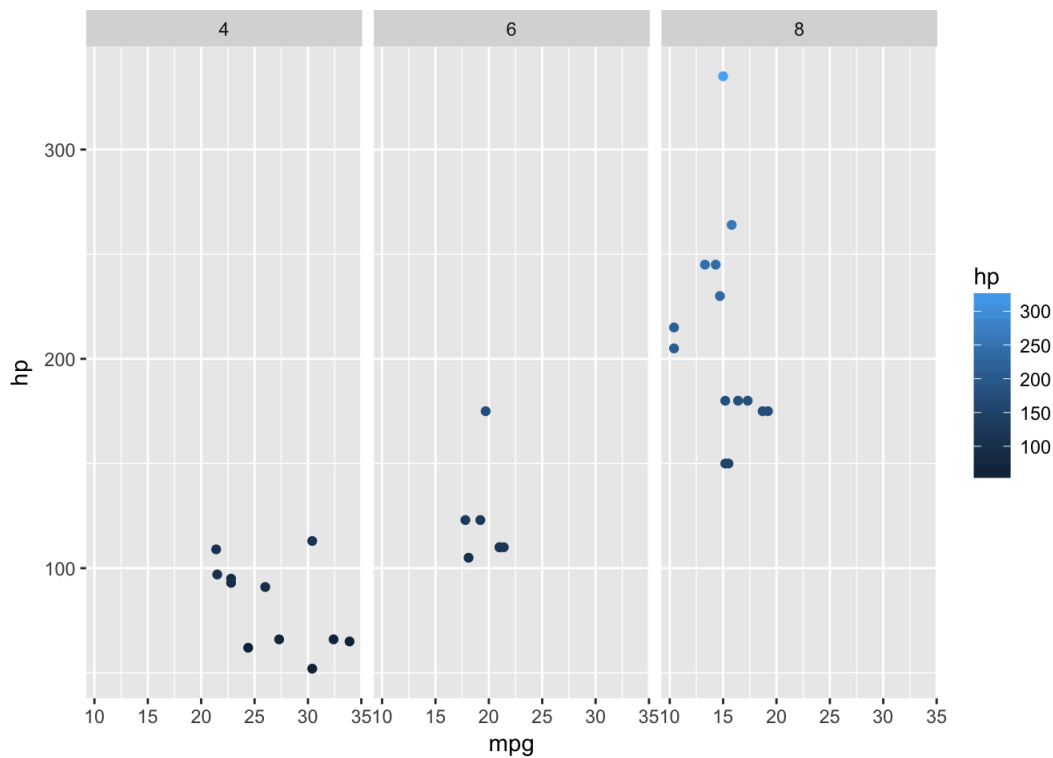


3. A scatterplot of `mpg` modeled by horsepower (`hp`). Create facets by the number of cylinders (`cyl`) and `gear`.

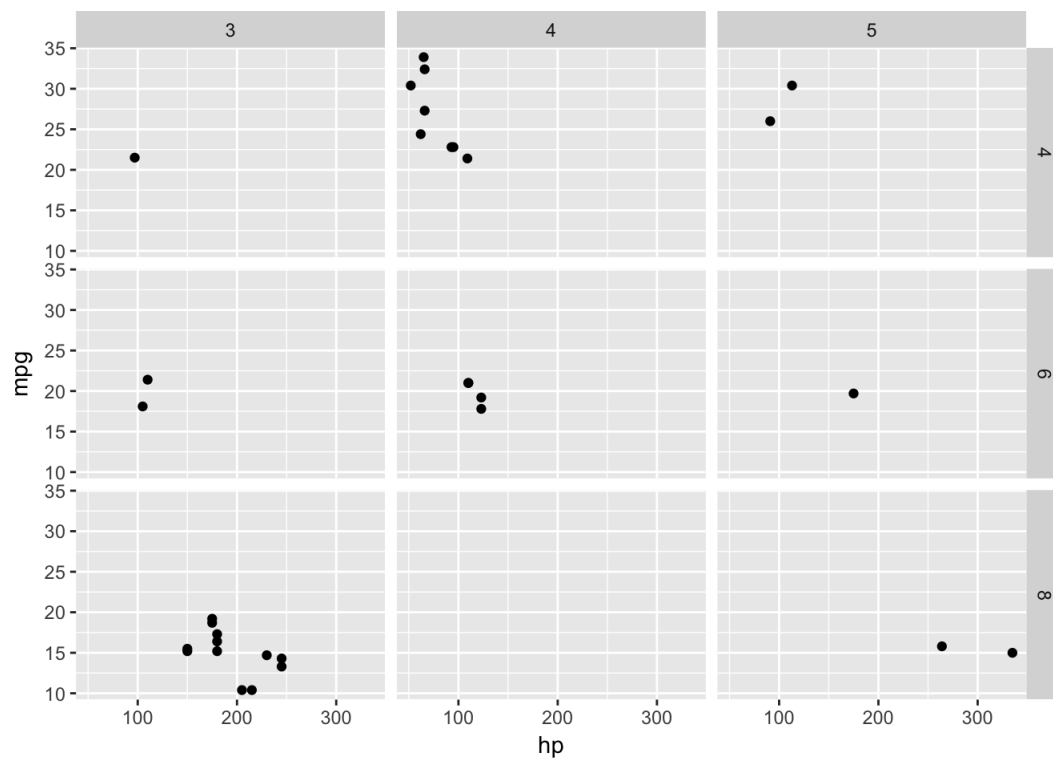
```
# Your code here
library(ggplot2)
View(mtcars)
```

```
## Error in check_for_XQuartz(): X11 library is missing: install XQuartz from xquartz.macosforge.org
```

```
ggplot(mtcars, aes(x = mpg, y = hp, color = hp)) + geom_point() + facet_grid(. ~
  cyl)
```



```
p1 <- ggplot(mtcars, aes(x = hp, y = mpg))
p1 + geom_point() + facet_grid(cyl ~ gear)
```



```
# p1 <- ggplot(Cars93, aes(x = Price)) p1 + geom_histogram() +
# facet_grid(Origin ~ DriveTrain)
```

Problem 3

Using the `batting` data set (UsingR), create a visualization that does the following:

- Plots the rate of intentional walks (that is, the number of intentional walks divided by the number of times a player was at bat; these are the `IBB` and `AB` variables in the data set, respectively) against the rate of home runs (the `HR` variable in the data set) as a scatterplot
- Draws a trend line for these variables
- Identifies and labels the outlier in the data set in these variables (easily spotted once the scatter plot is drawn)

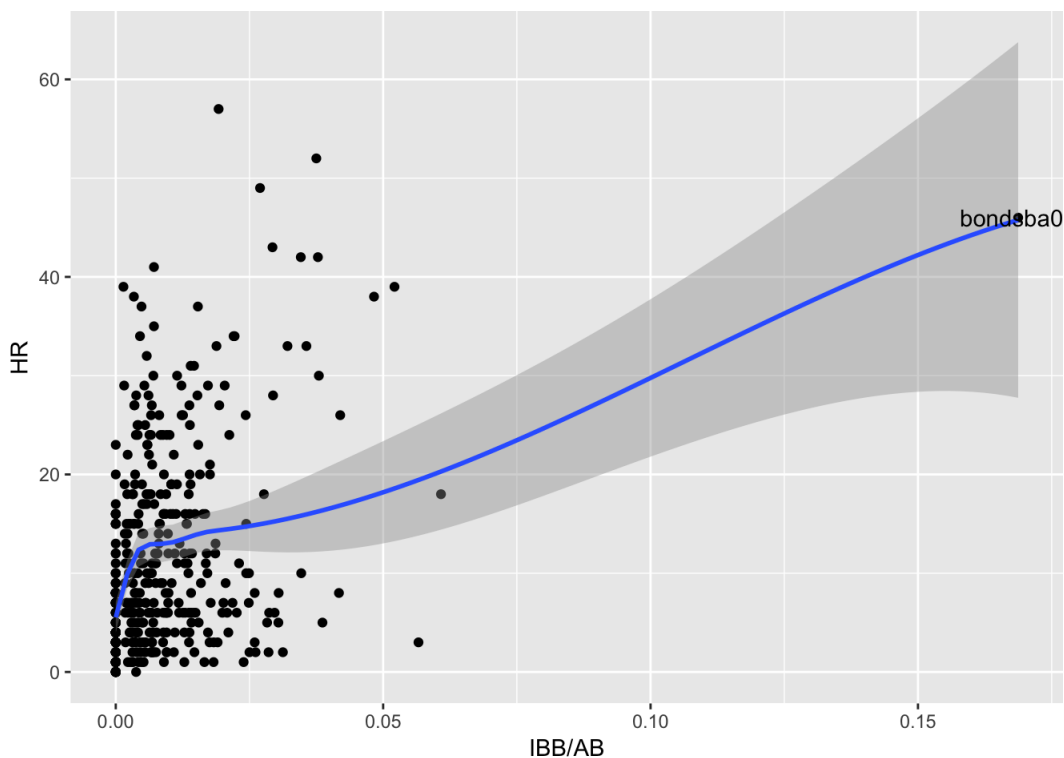
(Hint: `geom`-type functions can accept data arguments and will use the data set passed rather than the default for the chart. So for the third requirement, consider adding a text layer with `geom_text(data = ..., aes(...))` where the argument passed to `data` is a subset of the data set consisting of the outlier, and `aes(...)` defines how to label that outlier.)

```
# Your code here
library(UsingR)
View(batting)
```

```
## Error in check_for_XQuartz(): X11 library is missing: install XQuartz from xquartz.macosforge.org
```

```
p1 <- ggplot(batting, aes(x = IBB/AB, y = HR))
p2 <- p1 + geom_point() + stat_smooth()
p2 + geom_text(data = subset(batting, subset = IBB/AB > 0.1), aes(x = IBB/AB,
  y = HR, label = playerID))
```

```
## `geom_smooth()` using method = 'loess'
```



BONUS Problem

Reconsider the data set from a previous project containing data about the results of 2012 Olympics. I load the data in for you below:

```
olympic2012 <- read.csv("http://introcs.cs.princeton.edu/java/data/olympic-medals2012.csv")
# Some variables are read in as strings when, in truth, they are numeric (they separate thousands with commas, leading to them being read as strings). I fix this below using the transform function, which allows for modifying columns in a data frame using methods similar to with.
olympic2012 <- transform(olympic2012, GDP.2011 = as.numeric(gsub(",", "", GDP.2011)), pop.2010 = as.numeric(gsub(",", "", pop.2010)))
```

Use any plotting system (base R, **lattice**, **ggplot2**) to create plot involving at least three variables in the `olympic2012` data set. Explain the relationship you explored and any interesting findings. **Bonus points will be given for plots that I consider exceptionally clean, clear, and insightful, accompanied with good analyses of what you found.**

```
# Your code here
```