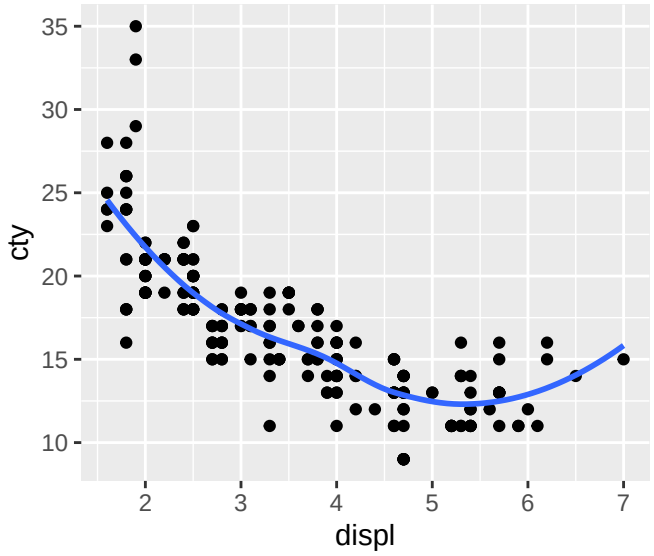
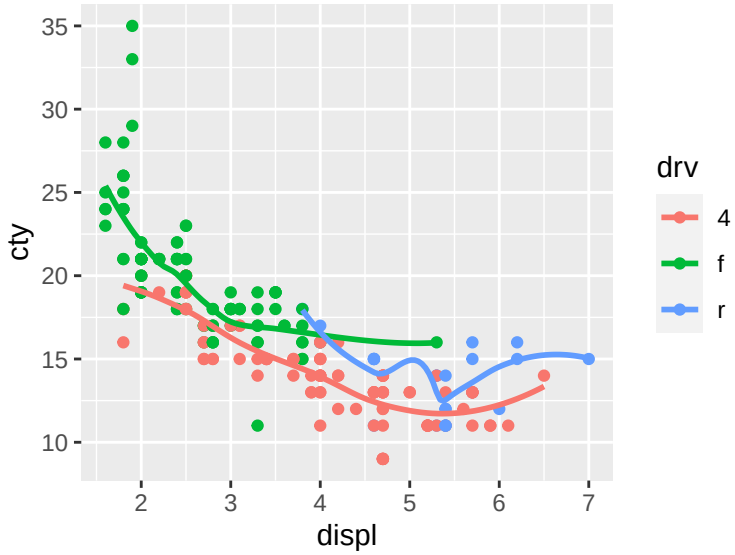
HW 2, STAT 450
Due: Thursday, September 30

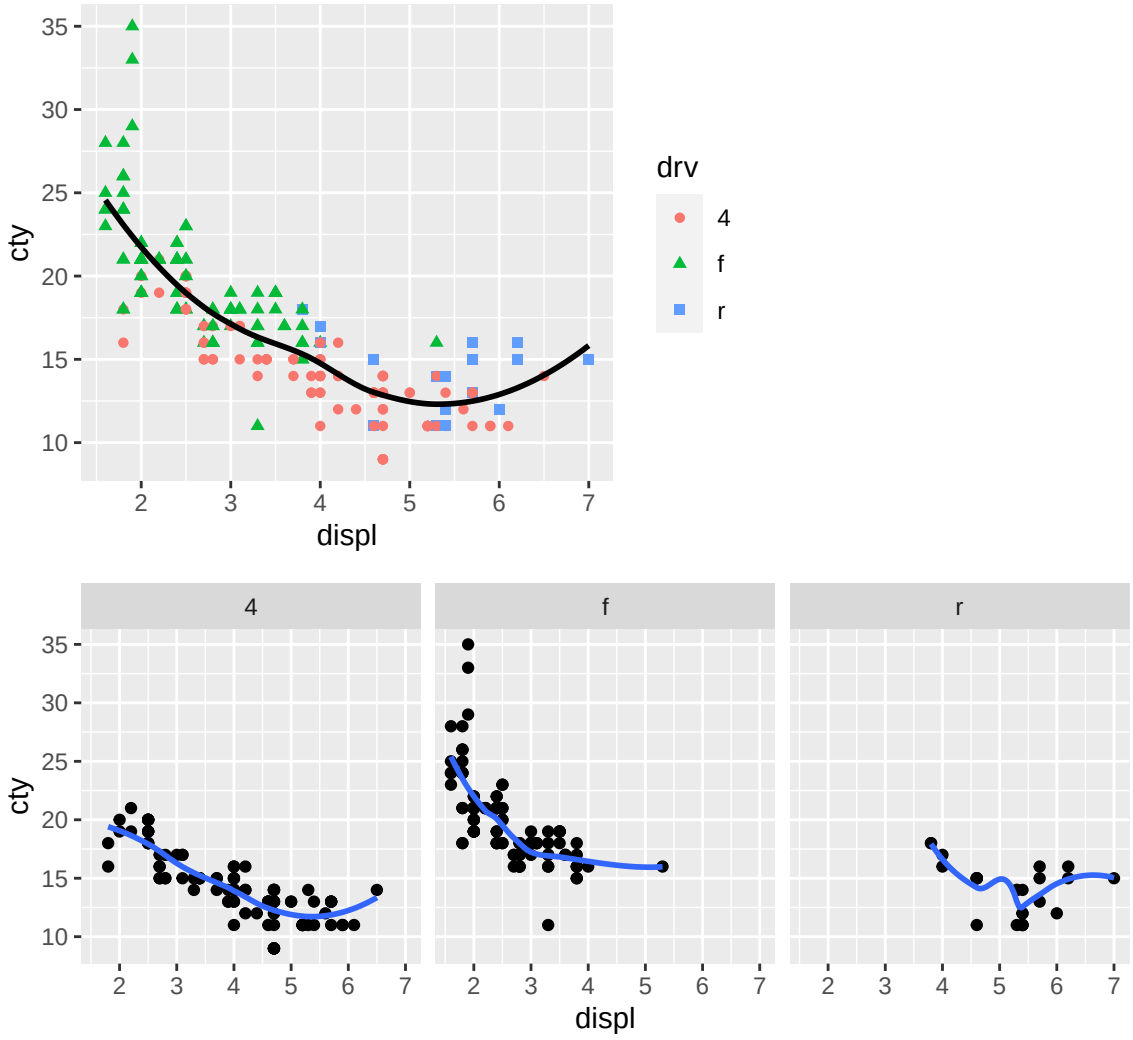
Reading: Chapters 3 of `https://r4ds.had.co.nz/`

Directions: Please submit your completed assignment to Blackboard. The assignment should be completed using R Markdown and rendered to an PDF or HTML format. Note that Blackboard will not accept HTML files. One workaround is to first zip your HTML file, and then submit the zipped file to Blackboard.

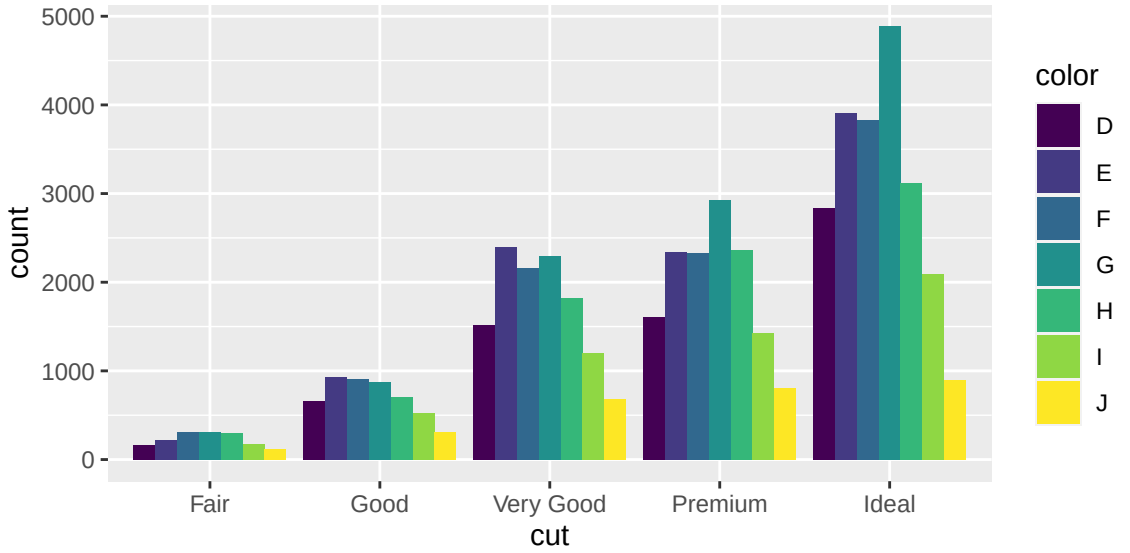
Exercise 1. Using the `mpg` data frame, recreate the R code necessary to generate the following graphs. In your submission, show both the R code and the graphs.

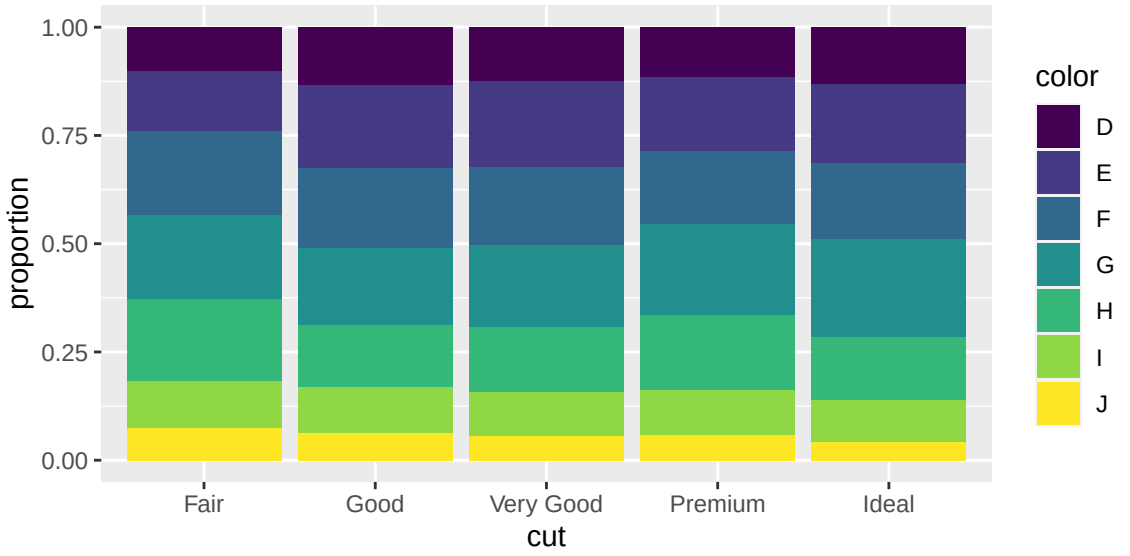





cty
displ
drv
4
f
r
4
f
r

Exercise 2. Using the `diamonds` data frame, recreate the R code necessary to generate the following graphs. In your submission, show both the R code and the graphs. Which diamond color is best, and which color is worst? (Hint: type `help(diamonds)` to read the documentation for this data set)





Exercise 3. In this exercise you will make a map of Alameda County. First, make sure to load the relevant map packages:

```
library(maps)
library(mapproj)
```

(a) Run the following code to make a map of California with county boundaries.

```
ca <- map_data("county", "ca")
ggplot(ca, aes(long, lat, group = group)) +
  geom_polygon(fill = "white", color = "black") +
  coord_map()
```

(b) The object `ca` is a data frame that contains the coordinates for the polygons of each county in California. Here is a preview of the first several rows:

```
head(ca)
```

```
##        long      lat group order     region subregion
## 1 -121.4785 37.48290     1     1 california   alameda
## 2 -121.5129 37.48290     1     2 california   alameda
## 3 -121.8853 37.48290     1     3 california   alameda
## 4 -121.8968 37.46571     1     4 california   alameda
## 5 -121.9254 37.45998     1     5 california   alameda
## 6 -121.9483 37.47717     1     6 california   alameda
```

Run the following two commands, and explain what you think each command is doing.

```
unique(ca$subregion)
length(unique(ca$subregion))
```

(c) Use the `dplyr` function `filter()` to extract the rows of the `ca` data frame that correspond to Alameda County. Store the subset in a new data frame called `alameda_ca`.

(d) Use the subsetted data frame from part c to make a map of Alameda County with `ggplot2`.

Bonus [2 points]: Make a map of the nine counties in the Bay Area (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma).