Homework #2

1. Flowers

Data: `flowers` dataset in **cluster** package

(a) Rename the column names and recode the levels of categorical variables to descriptive names. For example, "V1" should be renamed "winters" and the levels to "no" or "yes". Display the full dataset.

(b) Create frequency bar charts for the `color` and `soil` variables, using best practices for the order of the bars.

2. Minneapolis

Data: `MplsDemo` dataset in **carData** package

(a) Create a Cleveland dot plot showing estimated median household income by neighborhood.

(b) Create a Cleveland dot plot *with multiple dots* to show percentage of 1) foreign born, 2) earning less than twice the poverty level, and 3) with a college degree *by neighborhood.* Each of these three continuous variables should appear in a different color. Data should be sorted by college degree.

(c) What patterns do you observe? What neighborhoods do not appear to follow these patterns?

3. Taxis

Data: NYC yellow cab rides in June 2018, available here:

http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

It's a large file so work with a reasonably-sized random subset of the data.

Draw four scatterplots of `tip_amount` vs. `far_amount` with the following variations:

(a) Points with alpha blending

(b) Points with alpha blending + density estimate contour lines

(c) Hexagonal heatmap of bin counts

(d) Square heatmap of bin counts

For all, adjust parameters to the levels that provide the best views of the data.

(e) Describe noteworthy features of the data, using the "Movie ratings" example on page 82 (last page of Section 5.3) as a guide.

4. Olive Oil

Data: `olives` dataset in **extracat** package

(a) Draw a scatterplot matrix of the eight continuous variables. Which pairs of variables are strongly positively associated and which are strongly negatively associated?

(b) Color the points by region. What do you observe?

5. Wine

Data: `wine` dataset in **pgmm** package

(Recode the `Type` variable to descriptive names.)

(a) Use parallel coordinate plots to explore how the variables separate the wines by `Type`. Present the version that you find to be most informative. You do not need to include all of the variables.

(b) Explain what you discovered.