

# Homework 1

## *Data Science I*

Use the R package ggplot to make the following plots. Some general guidelines for making the plots are:

- No axis titles, legend title, or facet titles should be named by the default variable names
- All plots should have a title
- When relevant, units for variables should be provided in the plot

The problems for this homework assignment have been adapted from your course textbook ‘Modern Data Science with R’.

1. Using the RailTrail data set from the mosaic package:

```
library(mosaic)
head(RailTrail)
```

```
##   hightemp lowtemp avgtemp spring summer fall cloudcover precip volume
## 1      83     50   66.5      0      1    0        7.6    0.00    501
## 2      73     49   61.0      0      1    0        6.3    0.29    419
## 3      74     52   63.0      1      0    0        7.5    0.32    397
## 4      95     61   78.0      0      1    0        2.6    0.00    385
## 5      44     52   48.0      1      0    0       10.0    0.14    200
## 6      69     54   61.5      1      0    0        6.6    0.02    375
##   weekday dayType
## 1    TRUE  weekday
## 2    TRUE  weekday
## 3    TRUE  weekday
## 4   FALSE weekend
## 5    TRUE  weekday
## 6    TRUE  weekday
```

- a. Create a scatterplot of the number of crossings per day (volume) against the high temperature that day. Color the points in the plot by the amount of precipitation.
- b. Separate your plot into facets by weekday.
- c. Add regression lines to the two facets.
- d. Write an interpretation of what you see in the three plots.

2. The following questions use the Marriage data set from the mosaic package.

```
library(mosaic)
head(Marriage)
```

```
##   bookpageID  appdate ceremonydate delay  officialTitle person  dob
## 1   B230p539 10/29/96   11/9/96     11   CIRCUIT JUDGE  Groom  4/11/64
## 2   B230p677 11/12/96   11/12/96     0 MARRIAGE OFFICIAL Groom  8/6/64
## 3   B230p766 11/19/96   11/27/96     8 MARRIAGE OFFICIAL Groom  2/20/62
## 4   B230p892 12/2/96    12/7/96     5      MINISTER  Groom  5/20/56
## 5   B230p994 12/9/96    12/14/96     5      MINISTER  Groom 12/14/66
## 6  B230p1209 12/26/96   12/26/96     0 MARRIAGE OFFICIAL Groom  2/21/70
##      age      race prevcount prevconc hs college dayOfBirth      sign
```

```
## 1 32.60274    White      0    <NA> 12      7      102.0      Aries
## 2 32.29041    White      1 Divorce 12      0      219.0      Leo
## 3 34.79178 Hispanic     1 Divorce 12      3       51.5      Pisces
## 4 40.57808    Black      1 Divorce 12      4      141.0      Gemini
## 5 30.02192    White      0    <NA> 12      0      348.5 Saggitarius
## 6 26.86301    White      1    <NA> 12      0       52.5      Pisces
```

- a. Create an informative and meaningful data graphic.
- b. Identify each of the visual ques that you are using and describe how they are related to each other.

3. The macleish package contains weather data collected every ten minutes in 2015 from two weather stations in Whatley, MA. Create a data graphic that displays the average temperature over each 10-minute interval (temperature) as a function of time (when). Add a smooth trend line to the plot and change the color of the line from the default color.

```
library(macleish)
```

```
## Loading required package: etl
```

```
head(whately_2015)
```

```
## # A tibble: 6 x 8
##   when                temperature wind_speed wind_dir rel_humidity pressure
##   <dtm>                <dbl>         <dbl>   <dbl>      <dbl>      <int>
## 1 2015-01-01 00:00:00      -9.32          1.40    225.        54.6        985
## 2 2015-01-01 00:10:00      -9.46          1.51    248.        55.4        985
## 3 2015-01-01 00:20:00      -9.44          1.62    258.        56.2        985
## 4 2015-01-01 00:30:00      -9.3          1.14    244.        56.4        985
## 5 2015-01-01 00:40:00      -9.32          1.22    238.        56.9        984
## 6 2015-01-01 00:50:00      -9.34          1.09    242.        57.2        984
## # ... with 2 more variables: solar_radiation <dbl>, rainfall <int>
```

4. a. Using data from the nasaweather packages, use the geom\_path() function to plot the path of each tropical storm in the storms data table. Use color to distinguish the storms from one another and use faceting to plot each year in its own panel. Remove the legend with the names of the storms.
- b. Plot the storm 'Erin' on a map of the United States. Color the path of the storm by the wind speed at each point. Note that we did not learn how to plot maps in class so you will need to use Google to figure this out!