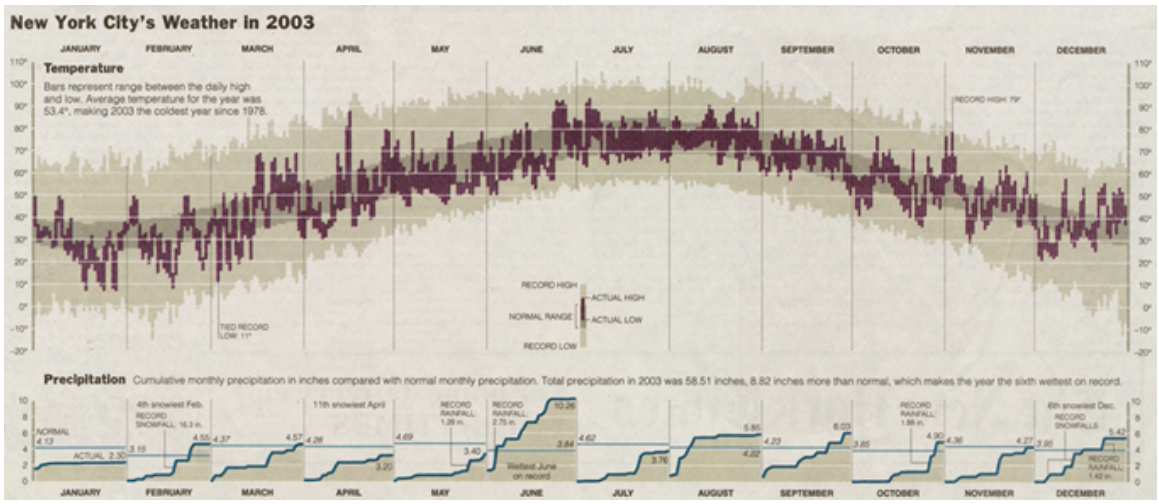
HW9 – Intro to Data Science
Due 9am Tuesday Nov 15 in Moodle (code only)
Hard copy due in class on Tuesday Nov 15 (including a printout of the graphic)

Often an effective approach to learning how to do something is to imitate a master. In this project you will recreate what many consider a master plot – one that provides a visualization of the daily weather in New York City for 2003 that appeared in the New York Times 4-Jan-2004.



It consists of two plots juxtaposed, one above the other. The top plot provides the daily high and low temperatures, the normal high and low, and record high and low. The bottom plot consists of 12 mini-plots, one for each month, providing the cumulative precipitation for the month. Larger versions of the plot are included at the end of this document. These plots are from
http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00014g

You are to create this same plot but with data from 2011 for San Francisco or Los Angeles. If your SID ends in an even number, you are to use the San Francisco data, if it ends with an odd number then you are to use the Los Angeles data. These data are available in moodle weather2011.rda

These data were retrieved from the National Weather Service Web Site, e.g., the weather for January 2011 in LAX can be found at:
http://www.wrh.noaa.gov/climate/monthdisp.php?stn=KLAX&p=temperature&mon=1&wfo=lox&year=2011

Below is a screenshot of one of the tables at the NOAA site.

KSFO July, 2011										
Date	Observed Low (F)	Observed High (F)	Normal Low (F)	Normal High (F)	Record Low (F)	Year	Record High (F)	Year	Observed Precipitation (inches)	Record Precipitation (inches)
1	54	80	54	71	46	1949	99	1985	0	.01 1980
2	53	70	54	71	47	1954	97	1970	0	.35 1977
3	56	84	54	71	48	1953	89	1970	0	.01 1955
4	55	69	54	71	46	1954	85	1973	0	0
5	52	78	54	71	48	1965	81	1981	0	0
6	52	70	54	71	47	1953	86	1957	0	0
7	53	69	54	71	49	1955	85	1985	0	0
8	53	72	54	71	48	1951	90	1985	0	.23 1974
9	52	66	54	71	47	1953	94	1959	0	.01 1945
10	53	69	54	71	47	1953	97	1959	0	0
11	53	63	54	71	48	1998	93	1983	T	0
12	54	65	54	71	47	1950	88	1983	0	0
13	55	61	54	71	49	1981	97	1972	T	0
14	55	63	54	71	48	1950	98	1972	0	0
15	55	64	54	71	49	1962	93	1995	T	.13 1975
16	56	70	55	71	49	1949	87	1988	0	.01 1990

You will work with the data frames, sfoWeather, sfoMonthlyPrecip, laxWeather, and laxMonthlyPrecip. The sfoWeather and laxWeather data frames contain the reported weather at the SFO and LAX airports. These data frames are identical in structure. The variables are described below:

Month: values 1 through 12 for January through December
Day: 1 – 31 for the day of the month
Low: observed low temperature for the day
High: observe high temperature for the day
NormalLow: normal low temperature for the day (based on a 30 year average)
NormalHigh: normal high temperature for the day (based on a 30 year average)
RecordLow: record low temperature for the day
LowYr: year in which the record low occurred
RecordHigh: record high temperature for the day
HiYr: year in which the record high occurred
Precip: observed precipitation for the day. The original table includes "T" for trace amounts of precipitation. These values have been converted to 0 in the data frame
RecordPrecip: record amount of precipitation for the day
PrecipYr: year in which the record precipitation occurred

The other two data frames (sfoMonthlyPrecip and laxMonthlyPrecip) each have 12 rows and two columns. Each row contains the total precipitation for the month and the total normal precipitation.

To help you, we provide some observations about the two plots.

In the top plot:
Four colors are used for the background, record temperature, normal temperature, and daily temperature.
A vertical dotted reference line appears for each month
Reference text is included for those days that tied or exceeded the record high/low.
The y-axis tick marks are at 10-degree intervals.
There is no x-axis, axis label, or tick marks, tick mark labels along the bottom of the plot. Instead, the months are included along the top axis.
There is a legend that indicates how to read the actual, normal, and record temperatures.
Each high and low for each day (whether actual, normal, record) is represented as a narrow rectangle.

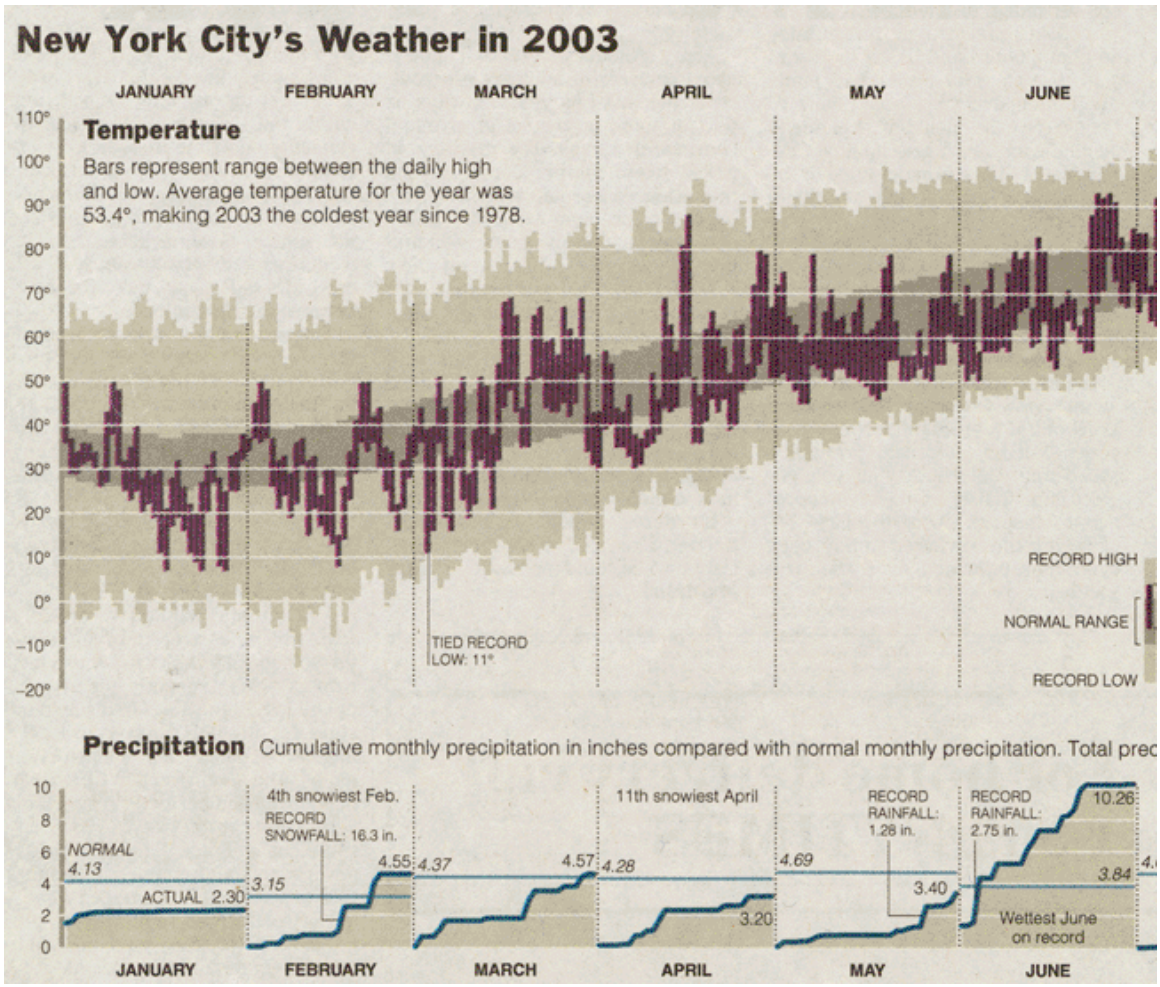
In the bottom plot:
There are 12 small plots, each provides a cumulative precipitation for each month.
Each has a reference line for the normal precipitation.
The area below the cumulative curve is shaded.
The x-axis is the mirror of the top plot.

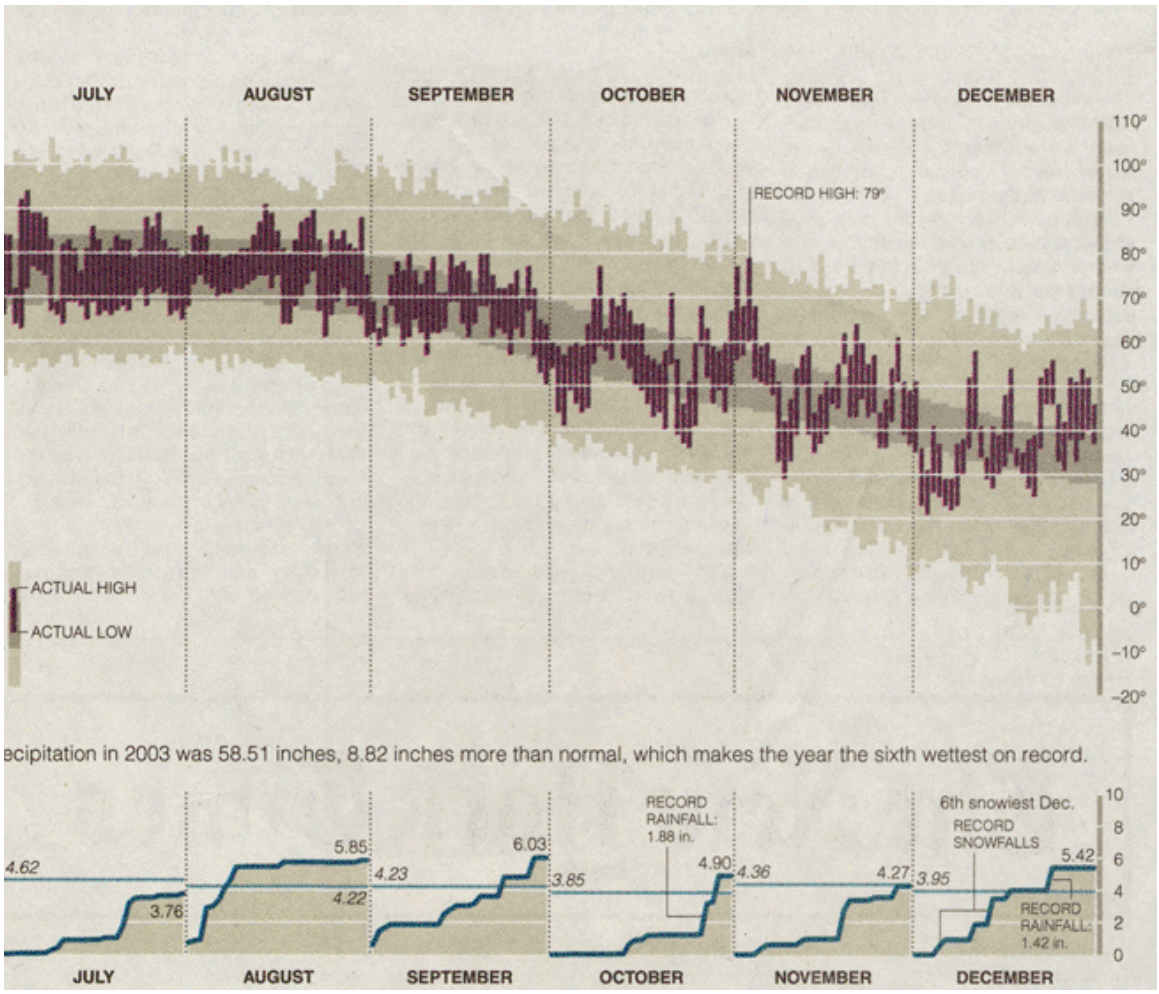
We have provided a .R file on Moodle with a structure for creating the plots.

2. Choose one of the Research LifeCycles we either discussed in class or is in the readings on moodle and explain in one or two paragraphs how your plot creation from this homework is described (or not described) by the lifecycle of research you chose.

EXTRA CREDIT: Design a plot that facilitates a comparison of San Francisco and Los Angeles weather.

MORE EXTRA CREDIT: Construct the plot for 2015.


New York City's Weather in 2003
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
110°
100°
90°
80°
70°
60°
50°
40°
30°
20°
10°
0°
−10°
−20°
Temperature
Bars represent range between the daily high and low. Average temperature for the year was 53.4°, making 2003 the coldest year since 1978.
TIED RECORD LOW: 11°
RECORD HIGH
NORMAL RANGE
RECORD LOW
Precipitation
Cumulative monthly precipitation in inches compared with normal monthly precipitation. Total prec
10
8
6
4
2
0
NORMAL 4.13
ACTUAL 2.30
4th snowiest Feb.
RECORD SNOWFALL: 16.3 in.
3.15
4.55
4.37
4.57
11th snowiest April
4.28
3.20
4.69
RECORD RAINFALL: 1.26 in.
3.40
RECORD RAINFALL: 2.75 in.
10.26
3.84
Wettest June on record
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE



ecipitation in 2003 was 58.51 inches, 8.82 inches more than normal, which makes the year the sixth wettest on record.