**Principles of Analytic Graphics**

* **Principle 1: Show comparisons**
  + - Evidence for a hypothesis
    - Always ask “Compare to what?”
* **Principle 2: Show causality, meachanism, explanation, systematic structure**
  + - What’s the causal framework to the current question?
* **Principle 3: Show multivariate data**
  + - Multivariate: > 2 variables
    - The real world is > 2 🡺 need to take them into consideration
* **Principle 4: Integration of evidence**
  + - Completely integrate words, numbers, images, diagrams
    - Data graphics should make use of many modes of data presentation
    - DO NOT let the tools control the analysis (*Which I will*)
* **Principle 5: Describe and document the evidence with labels, scales, sources**
* **Principle 6: CONTENT**
  + - There should be a good story to tell, do not go too vague

**Exploratory Graphs**

* **Why graphs?**
  + - Understand data properties
    - Find patterns
    - Suggest modeling strategies
    - “debug” analyses
    - Communicate results
* **Characteristic of exploratory graphs?**
  + - Made quickly
    - Large number are made
    - For personal understanding
    - Axes/legends are generally cleaned up
    - Color/size are just used for information

**Simple summaries of Data**

* **Five number summary**

Summary(dataset)

* **Boxplots**

Boxplot(dataset)

abline(h = 12)

* **Histograms**

Histogram(dataset, breaks = 100)

* **Density plot**

Hist(dataset)

* **Barplot**

Barplot(table(dataset), col = “wheat”, main = “name”)

**Simple summaries of Data**

* **Two dimensions** 
  + - Multiple/overlayed 1-D plot (lattice/ggplot2)
    - Scatterplots
    - Smooth scatter plots
* **> 2 dimensions**
  + - Overlayed/multiple 2D plots, coplots
    - Color, size, shape to add dimensions
    - Spinning plots
    - Actual 3D plots

**Multiple boxplots**

Boxplot(pm25 ~ region, data = pollution, col = “red”)

**Scatterplot**

With(pollution, plot(latitude, pm25, col = region))

Abline(h = 12, lwd = 2, lty = 2)

**Plotting systems in R**

* **3 core plotting systems** 
  + - **The base plotting system** 
      * Start with plot()
      * Use annotations functions to modify: text, lines, points, axis

library(datasets)

data(cars)

with(cars, plot(speed, dist))

* + - **The lattice system**
      * In the lattice package
      * Start with xyplot(), bwplot()
      * Good for putting many many plots on a screen

Library(lattice)

State <- data.frame(state.x77, region = state.region)

Xyplot(Life.Exp ~ Income | region, data = state, layout = c(4, 1))

* + - **The ggplot2 system**
      * Splits the difference between base and lattice
      * Automatically deals with spacings, text, titles
      * Superficial similarity to lattice

library(ggplot2)

data(mpg)

qplot(displ, hwy, data = mpg)

**Base plotting system in R**

* Core Plotting system
  + - Graphics: including plot, hist, boxplot, etc.
    - grDevices: all the code implementing graphics devices: X11, PDF, PostScript, PNG, etc.
* lattice uses
  + - lattice: including xyplot, bwplot, levelplot
    - grid: the lattice package is built on top of this shit

**The process of making a plot**

* Where will the plot be made? On the screen? In a file?
* How will it be used?
  + - Is the plot for viewing temporarily on the screen?
    - Will it be presented in a web browser?
    - Will it need to be printed on paper?
    - Will it be used in a presentation?
* How big is the data on the plot? Large set of few points?
* Will you need to resize it dynamically?
* What graphics system to be used? Base, lattice, ggplot2? These cannot be mixed

**Base graphics**

* 2 phases to create a base plot
  + - Initializing a new plot
    - Annotating an existing plot
* Calling plot(x, y) or hist(x) draw a new plot on the console
* Check ?par to see the parameters of this

Library(datasets)

Hist(airquality$Ozone)

With(airquality, plot(Wind, Ozone)

Boxplot(Ozone ~ Month, airquality, xlab = “Month”, ylab = “Ozone (ppb)”)

* **Key parameters**
  + - Pch: the plotting symbol (default is a open circle)
    - Lty – the line type
    - Lwd – the line width, integer
    - Col – the plotting color. The colors() gives you a vector of colors by name
    - Xlab – label for the x axis
    - Ylab – label for the y axis
    - Las – the orientation of axis labels
    - bg – background color
    - Mar – margin size
    - Oma – outer margin size
    - Mfrow – number of plots per row
    - Mfcol – numbers of plots per column
* See **default value** for parameters

Par(“bg”)

**Base plotting functions**

* Plot – make a scatterplot
* Lines – add lines to a plot
* Points – add points to a plot
* Text – add text labels to a plot using specified x, y coordinates
* Title – add annotations to x, y axis labels, title, subtitle, outer margin
* mtext – add arbitrary text to the margins
* Axis – adding axis ticks/labels

**Base plot with Annotation**

With(airquality, plot(Wind, Ozone, main = “name”)

With(subset(airquality, Month == 5, points(Wind, Ozone, col = “blue”))

Legend(“topright”, pch = 1, col = c(“blue”, “red”), legend = c(“May”, “Other Months”))

**Base plot with Regression Line**

With(airquality, plot(Wind, Ozone, main = “name”, pch = 20))

Model <- lm(Ozone ~ Wind, airquality)

Abline(model, lwd = 2)

**Multiple Base Plots**

Par(Mfrow = c(1, 2))

With(airquality, {

Plot(Wind, Ozone, main = “Ozone and Wind”)

Plot(Solar.R, Ozone, main = “Ozone and Solar Radiation”)

Plot(Temp, Ozone, main = “Ozone and Temperature”)

Mtext(“name”, outer = TRUE)

})

**Graphics Devices in R**

* **What is a Graphic Device?**
  + - Something where you can make a plot appear
      * A window on your computer
      * A PDF file
      * A PNG/JPEG file
      * A scalable vector graphics (SVG)
    - When making a plot, it has to be sent to a specific graphics device
    - The most common place? – the screen device – windows()
* **?Devices**
* **How does a plot get created?**
  + - **1st way**
      * Call a plotting function plot, xyplot, qplot
      * The plot appears on the screen device
      * Annotate the plot if necessary
      * Done!
    - **2nd way – just for file devices**
      * Launch a graphics device
      * Call a plotting function to make a plot
      * Annotate plot if necessary
      * Explicitly close graphic device with dev.off()
* **Graphics file devices**
  + - **2 types – vector and bitmap devices**
    - **Vector formats**
      * Pdf
      * svg
      * win.metafile
      * Postscript
    - **Bitmap formats**
      * Png
      * Jpeg
      * Tiff
      * Bmp