INTRODUCTION TO R FOR DATA VISUALIZATION

ASUCRP/JRSA Conference
August 31, 2021
OVERVIEW OF COURSE

PART I
- Basic graphs
  - Histogram
    - Distributions
  - Bar chart
  - Scatterplot
    - Fitted line

PART II
- ggplot2
- Tables
- Copying/writing graphs/tables
  - LaTeX
- Brief intro to Tidyverse
Lindsay A. Bostwick, Ph.D.

- Research Associate, JRSA
- Specializations: Criminology, Sentencing and Corrections, Quantitative Analysis
- lbostwick@jrsa.org
setwd("C:\Users\Lindsay\Dropbox\First paper\First paper\Writeup\Figures")
diffs<-read.csv(file="Differences exit-admit.csv", header=T, sep=";")
str(diffs)
attach(diffs)
library(Hmisc)
win.metafile("C:\Users\Lindsay\Dropbox\First paper\First paper\Writeup\Figures\Figures for presentation\diffs2.wmf")
par(mfrow=c(1,1))
plot(c(1990,2012), c(-100,50), type="n", xlab="Year", ylab="Difference in months", main="Difference between exit and admission cohort estimates", family="serif")
lines(AdmitYear, Assault.diff, col="blue", pch=15, lwd=2, type="o", family="serif")
lines(AdmitYear, Arson.diff, col="cornflowerblue", pch=16, lwd=2, lty=2, type="o", family="serif")
lines(AdmitYear, Burglary.diff, col="green", pch=17, lwd=2, lty=3, type="o", family="serif")
lines(AdmitYear, Drugs.diff, col="darkorchid", pch=18, lwd=2, lty=5, type="o", family="serif")
lines(AdmitYear, DUI.diff, col="darkgoldenrod2", pch=19, lwd=2, lty=6, type="o", family="serif")
lines(AdmitYear, Robbery.diff, col="darkmagenta", pch=3, lwd=2, lty=1, type="o", family="serif")
lines(AdmitYear, Sex.diff, col="aquamarine4", pch=4, lwd=2, lty=2, type="o", family="serif")
lines(AdmitYear, weapons.diff, col="red", pch=8, lwd=2, lty=3, type="o", family="serif")
legend("bottomright", c("Assault", "Burglary", "Drugs", "DUI", "Robbery", "Sex Crimes", "Weapons"), lty=c(1,3,5,6,1,2,3), lwd=c(2,2,2,2,2,2,2), col=c("blue", "green", "darkorchid", "darkgoldenrod2", "darkmagenta", "aquamarine4", "red"), cex=0.8, pch=c(15,17,18,19,3,4,8))
minor.tick(ny=5, nx=0)
abline(0,0)
dev.off()
# AHH BEAUTIFUL!!

win.metafile("C:\Users\Lindsay\Dropbox\First paper\First paper\writeup\Figures\Figures for presentation\violent with bars2.wmf")

par(mfrow=c(2,2))
par(family="serif")

# Sex Crimes
plot(AdmitYear, Sex.Crimes, type="o", xlim=c(1983, 2012), ylim=c(0,220), xlab="Admission Year",
     ylab="Average time served in months", main="Average time served in months for Sex Crimes",
     col="blue", pch=19, lwd=2, lty=1, family="serif")

len=.07
for (i in 1:30) {
  arrows(AdmitYear[i], Sex.Crimes[i], AdmitYear[i], Sex.Crimes[i]+SexError[i], angle=90, length=len)
  arrows(AdmitYear[i], Sex.Crimes[i], AdmitYear[i], Sex.Crimes[i]-SexError[i], angle=90, length=len)
}

LABELS

- Main title
  - Main = "Main title here"
  - Multiple lines (\n):
    - main = "Main title\nhere"
- X-axis label
  - xlab="X-axis title here"
- Y-axis label
  - ylab="Y-axis title here"
- Typeface
  - par(family="serif")
- Defaults
* Not made by prior syntax (uses ggplot)
- PCH & BG
- CEX
- COL
- LWD & LTY & ARROWS
- ABLINE

COLORS, LINES, MARKERS
PCH & BG

Plot character (marker)

PCH = point type

BG = background color

**pch arguments**

- □ 0
- ○ 1
- △ 2
- ━ 3
- × 4
- ◊ 5
- ◄ 6
- □ 7
- ◦ 8
- ◊ 9
- ◄ 10
- □ 11
- □ 12
- □ 13
- □ 14
- ■ 15
- ■ 16
- △ 17
- ◊ 18
- ● 19
- ● 20
- ● 21
- △ 22
- ◊ 23
- △ 24
- ▼ 25
CEX

- CEX (magnify text outside of plot)
  - cex.main (main titles)
  - cex.axis (axes)
  - cex.lab (labels)
COL

- Colors
  - Defaults
    - `col="blue"`
  - RColorBrewer
    - `col="set2"`
LWD & LTY & ARROWS

- LWD = line width
  - Default = 1
  - 2 is twice as wide
- LTY = line type
- ARROWS = arrow type
**ABLINE**

- `abline()`
  - Adds lines to plot
  - Horizontal, vertical, regression
- `abline(intercept, slope)`
- `abline(h=0)` for horizontal line (y-value)
- `abline(v=0)` for vertical line (x-value)
- `abline(lm(y~x))` for fitted regression line

Graph:

![Graph of y = 2x + 1](image)
PART II: GGPLOT, Tables, & Tidyverse

QUESTIONS?
- Can be installed by itself
  - `install.packages("ggplot2")`
- Can be installed within `tidyverse`
  - `install.packages("tidyverse")`
TO START

- Begin with plot function
  - `ggplot()`
- Creates coordinate system to add layers to
- Define dataset to use
  - `ggplot(data=data)`
- Creates an empty graph
- Add layer of points to plot
  - `geom_point()`
- Mapping argument
  - Defines how variables in data are mapped to plot
  - Always include `aes()`
    - `aes` specifies which variables to map to x and y axes
Aesthetic mapping

- Visual property of objects in a plot
- E.g., size (size=), shape (shape=), color (color=), transparency (alpha=) of points
  - Note on shapes – ggplot will only use six shapes at a time
- Can map the aesthetics in plot to variables in data
  - Example: change colors in scatterplot based on a third variable
- When just changing, say, all the colors in a plot, set it outside the aes
  - ggplot(data=data) + geom_point(mapping=aes(x=, y=), color="red")
- Shapes of point is character; Size of point is mm; Shape of point is number
  - Border determined by “color” and fill determined by “fill”
- Legends are automatically created but can be suppressed
Can be one of the hardest aspects to wrap (hehe) your head around in ggplot
- Often need to play around and make mistakes
- Will cover the simplest versions here

Can split a plot into subplots or facets (particularly categorical variables)

Single variable: `facet_wrap()`
- `facet_wrap(~ categorical variable, nrow(#))`

What is `~`?
- The structure of the data
  - Another course

Facet grid
- Plots on the combination of two variables
  - `facet_grid(var1 ~ var2)`
  - `Facet_grid(.~var2)`
- Geom
  - Geometrical object used to represent the data
- geom_point
  - Scatterplots
- geom_smooth
  - Smooth line fitted to data
- Takes an aes function
- Again, must play with this for shape, size, linetype, etc
  - Some work with some geoms while others will not
- Can include two geoms on the same graph
  - Can have different ones for different layers
Many other geom functions such as:

- Title
- Jitter
- Scale
- Ribbon
- Theme
- Coordinate system
  - Hardest to learn
## Assault

```r
plot.dataaslt <- subset(plot.data, offtype=='Assault')
qaslt <- ggplot(plot.dataaslt, aes(x=Admityear)) + geom_line(aes(y=alos), linetype=1, size=1) +
    geom_ribbon(aes(ymin=lowerCI, ymax=upperCI), alpha=0.1, linetype=2) + theme_bw() + ylab('Average time served in months') +
    xlab('Admission year') + scale_y_continuous(limits = c(0, 115)) +
    ggtitle("Aggravated Assault") + theme(plot.title = element_text(face = 'bold'))
```

## Robbery

```r
plot.datarob <- subset(plot.data, offtype=='Robbery')
qrob <- ggplot(plot.datarob, aes(x=Admityear)) + geom_line(aes(y=alos), linetype=1, size=1) +
    geom_ribbon(aes(ymin=lowerCI, ymax=upperCI), alpha=0.1, linetype=2) + theme_bw() + ylab('Average time served in months') +
    xlab('Admission year') + scale_y_continuous(limits = c(0, 95)) +
    ggtitle("Robbery") + theme(plot.title = element_text(face = 'bold'))
```
- **sjPlot**
  - Good for making simple tables
    - Crosstabs
    - Regression results
    - Other tables
  - Can plot as ggplot figure or html (includes CSS-argument)
  - Uses a lot of ggplot functions/arguments
  - Tables can be hard to get out

- **apaTables**
  - Creates publishable tables in APA format
  - Still no beautiful way to get them out yet

**TABLES**
Multiple ways to do this

- `write.graph`
- `win.metafile graph code dev.off()`
- `pdf("LOCATION/name.pdf") graph code dev.off()`
- `png("LOCATION/name.png") graph code dev.off()`
- `jpeg("LOCATION/name.jpeg") graph code dev.off()`
- `tiff("LOCATION/name.tiff") graph code dev.off()`
New collection of packages

- [https://www.tidyverse.org/](https://www.tidyverse.org/)
- `ggplot2` – plotting and graphics
- `dplyr` – data manipulation
- `tidyr` – every variable a column, every column is variable
- `readr` – reads rectangular files like `.csv`
- `purr` – functions and vectors
- `tibble` – data frame manipulation and simplification of code
- `stringr` – working with string data
- `forcats` – working with factor variables
R SOFTWARE & PACKAGES

- https://www.r-project.org/
- Free
- Open-source
- Developer packages
- R Studio
  - https://www.rstudio.com/
- ggplot2
  - https://ggplot2.tidyverse.org/
- sjPlot
  - https://cran.r-project.org/web/packages/sjPlot/sjPlot.pdf
- tidyverse
  - https://www.tidyverse.org/
OTHER RESOURCES

- https://www.statmethods.net/
- RColorBrewer
- R Graph Gallery
- Plotting anything with ggplot2
  - https://www.youtube.com/watch?v=h29g21z0a68
  - https://www.youtube.com/watch?v=0m4yywqNPVY
BOOKS & WEB RESOURCES

- Cookbook for R
  - http://www.cookbook-r.com
  - https://rc2e.com/

- R Graphics Cookbook
  - Amazon link

- R for Data Science
  - https://r4ds.had.co.nz/
  - Amazon link

- R Studio Cheat Sheets
  - https://www.rstudio.com/resources/cheatsheets/