09 Graphics for Statistical Analysis

36-721 Statistical Graphics and Visualization

Jerzy Wieczorek

9/29/15

Last time

- Visually demonstrating statistical concepts & algorithms
- Shiny lab session
- Project 1 due

Today

- Critiques assigned, due on Thursday in class
- Graphics useful for statistical analysis

Follow along

- Editable code in 09_StatAnalysis_code.R
- Code with output examples in 09_StatAnalysis_code.html
- Data in tips.csv, ganglion.txt

Mathematical formulas and distributions

- `curve()` for plotting formulas
- Probability density functions (PDFs): `dnorm()` etc (d for density)
- Cumulative density functions (CDFs): `pnorm()` etc (p for probability)
- Contours of bivariate functions

Empirical distributions

- Histograms; choosing bins
- Kernel density estimates (KDEs); choosing kernel and bandwidth
- Empirical CDFs (ECDFs)
- Quantile-quantile (Q-Q) plots: do distributions match?
- Boxplots and bagplots
- Time maps

Precision / uncertainty of estimates

- Error bars; label whether they are confidence intervals or standard errors!
- Regression / smoother confidence bands

Regression models and diagnostics

Examples are based on Cleveland, *Visualizing Data*, Ch. 3.3 and 4.1

- ▶ RDA vs EDA (rote vs exploratory): checking your models leads to multiple comparisons, but *not* checking your models leads to terrible models
- ▶ Modeling decisions
  - ▶ Selecting model terms
  - ▶ Variable transformations
- ▶ Checking assumptions
  - ▶ No trend in residuals
  - ▶ No trend in spread of residuals
  - ▶ Residuals are approx Normal

More: examples using `lattice` for bivariate or trivariate data

Time series data, very briefly

- Time series data
- Autocorrelation
- Periodicity

More: examples using `lattice`

Other tools

R package `rgl` for draggable 3D plots of data, mathematical functions, and terrain

For next time (change from syllabus)

- Critique due Thurs. 3pm, in class
- Thursday 10/1: maps, map projections, principles of cartography, shapefiles in R
- Next Tuesday 10/6: **no class**
- Next Thursday 10/8: high-dimensional data; install GGobi to follow along