Why do we need statistics?

Course Introduction

Statistics is the science of collecting, analyzing, and drawing conclusions from data in a sensible way.

Statistics offers us powerful tools for gaining insight into the world around us.

"It is the mark of a truly intelligent person to be moved by statistics" - George Bernard Shaw

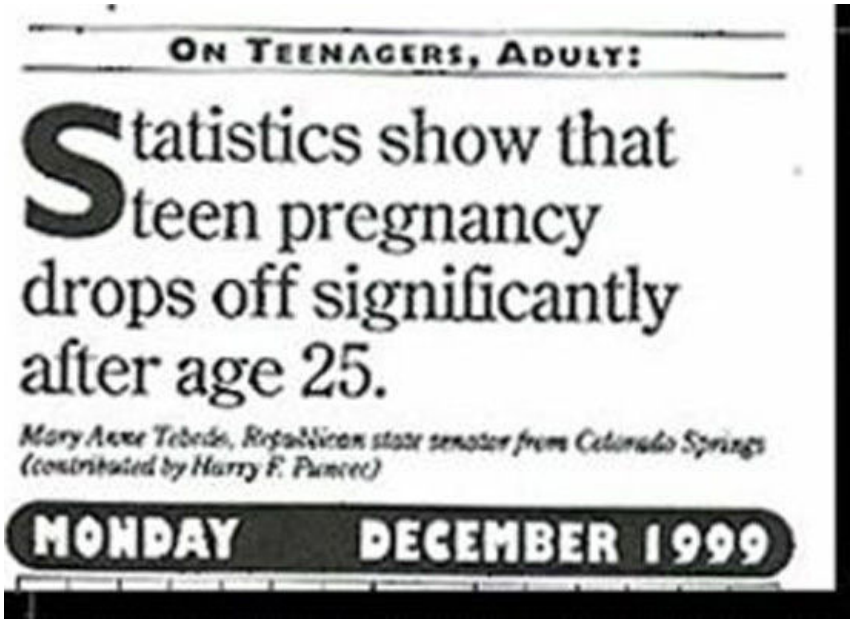
WHAT I AM GOING TO LEARN

my goals for the class

- Statistically literate
  - understand inference (e.g., study design)
  - know when something is causal and when it is just associated
  - know what a p-value indicates and what it doesn't
  - know the assumptions about basic tests and what the consequences are if they are ignored
- Appreciate visuals
  - see the value of graphs at the beginning of an analysis
  - see the value of graphs for disseminating your research
- Know when and where to seek help

Book Version

- Why are statistical procedures necessary at all?
- How can statistics help in planning experiments?
- Which procedure should I employ to analyze the results?
- What do the statistical results actually mean when I've got them?



THE IMPORTANCE OF STATISTICAL LITERACY

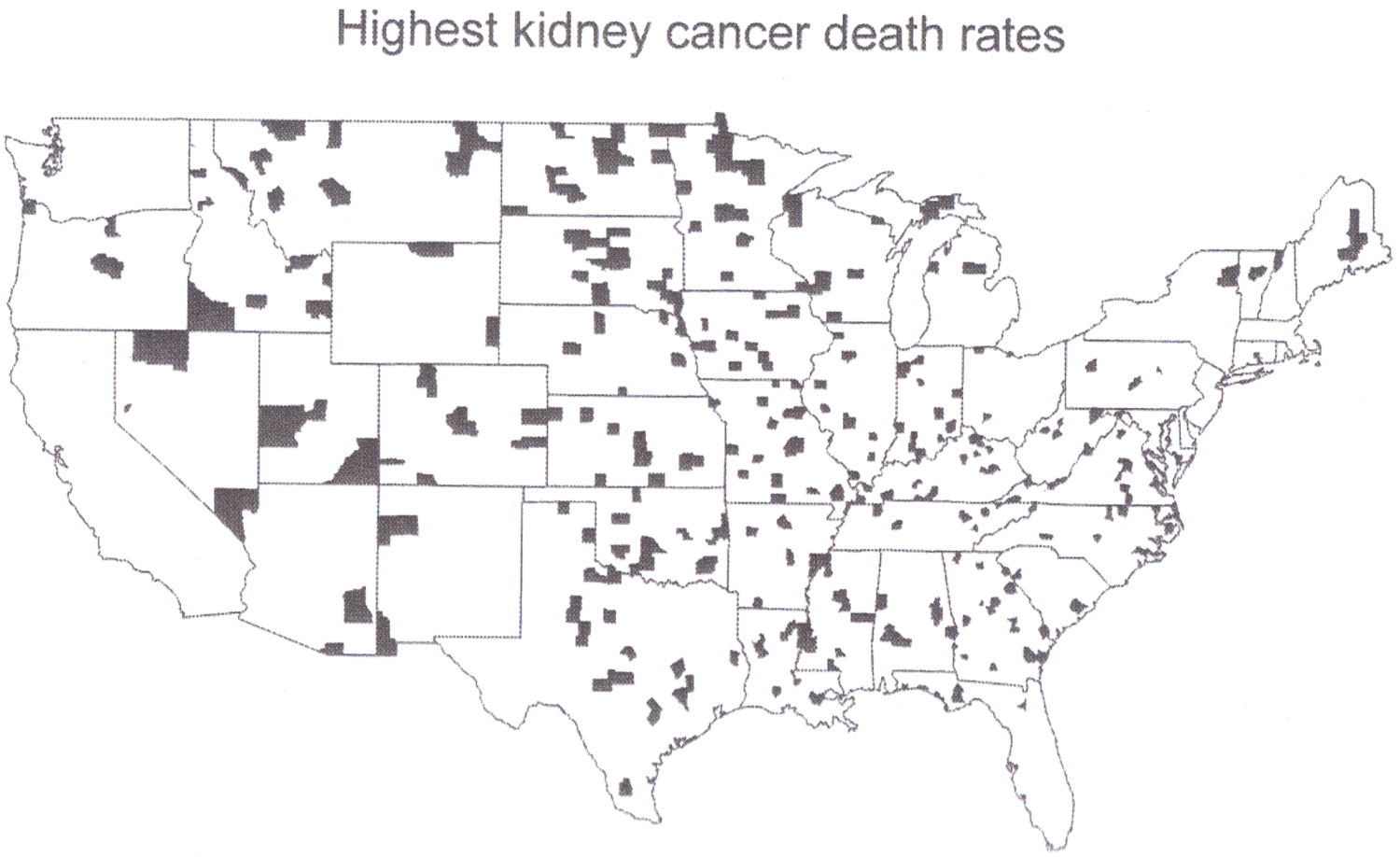
Headlines
(courtesy of http://www.statschat.org.nz/)

- Surprising Cancer Causes – stuff.co.nz

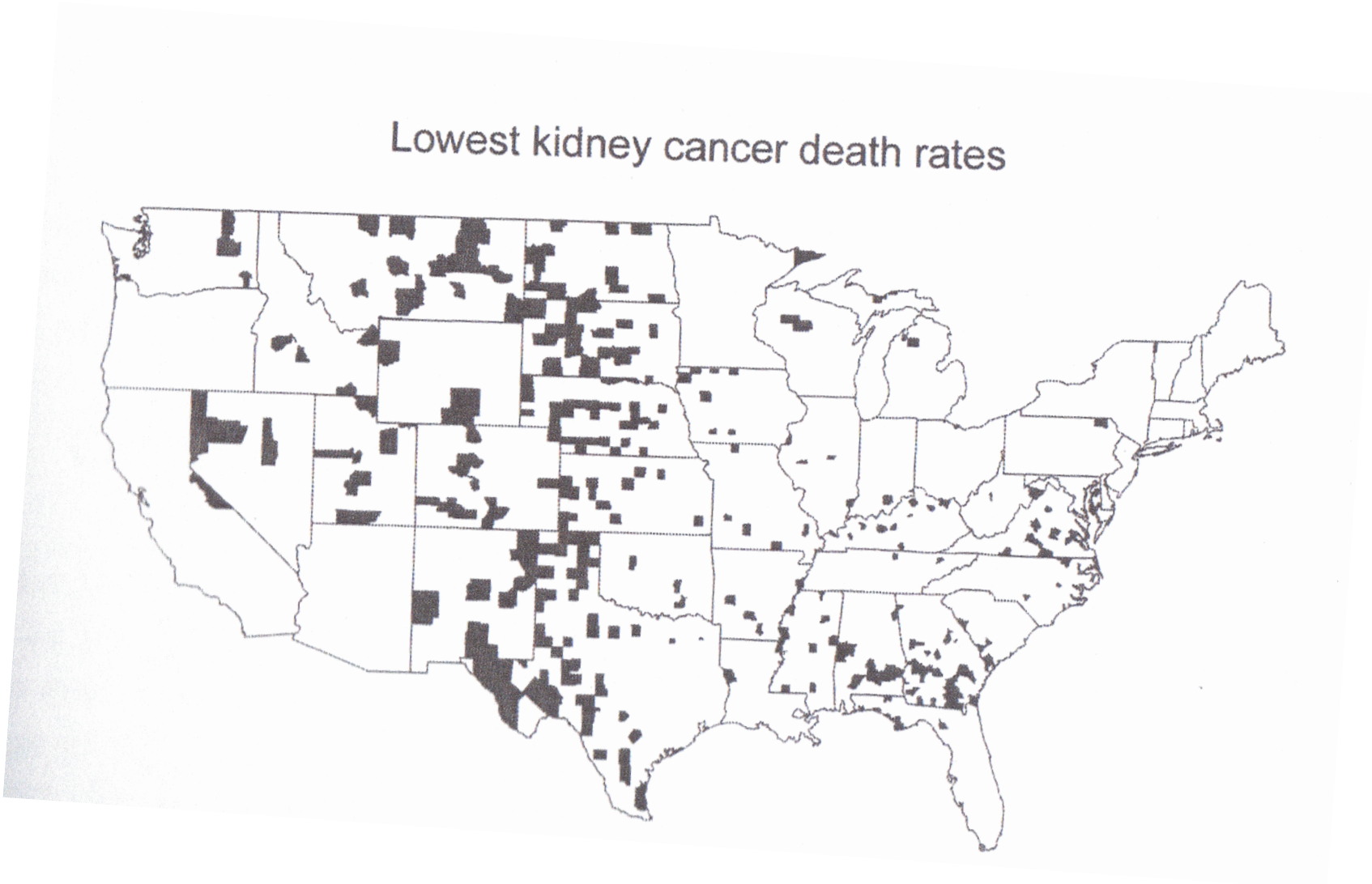
"There may be a darker reason a pet dog is a cancer patients' best friend. Analysis of breast cancer cases by researchers at the University of Munich showed that 79.7 per cent of all breast cancer patients had regular contact with dogs before diagnosis. Only 4.4 percent of the patients did not have pets at any time, compared to 57.3 percent of a healthy control group. According to researchers, that's a **29-fold** increased risk for pet owners. "

Fido and Cancer

- Lifetime risk of breast cancer – 10%
  - 29-fold increase?
- Actual relative risk reported in paper – 3.5
- Reported in "Medical Hypotheses"
  - self described as "intended as a forum for unconventional ideas without the traditional filter of scientific peer review"
- "*We compared the frequencies of dog and pet ownership with data from public available statistics on women (N=1320) of the same age group in Bavaria"*
  - Detail questionnaire about contact with dogs for subjects, but one simple question for controls.

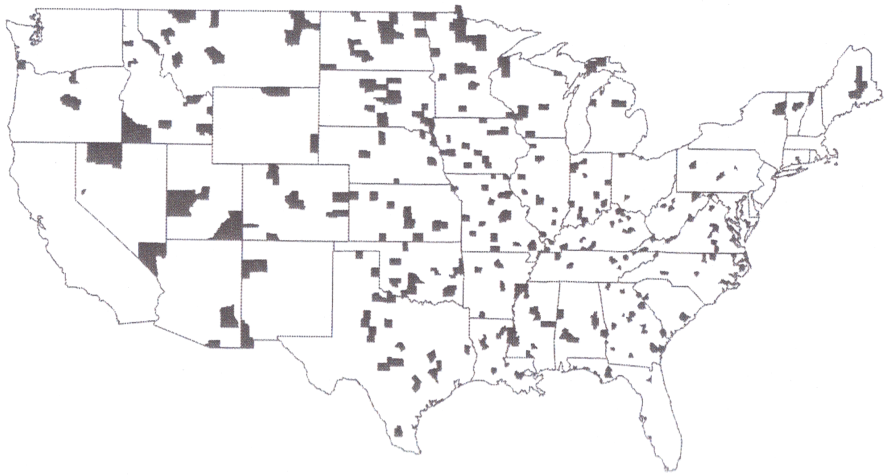


The counties of the US with the highest 10% age-standardized death rates for cancer of kidney of US white males, 1980-1989.

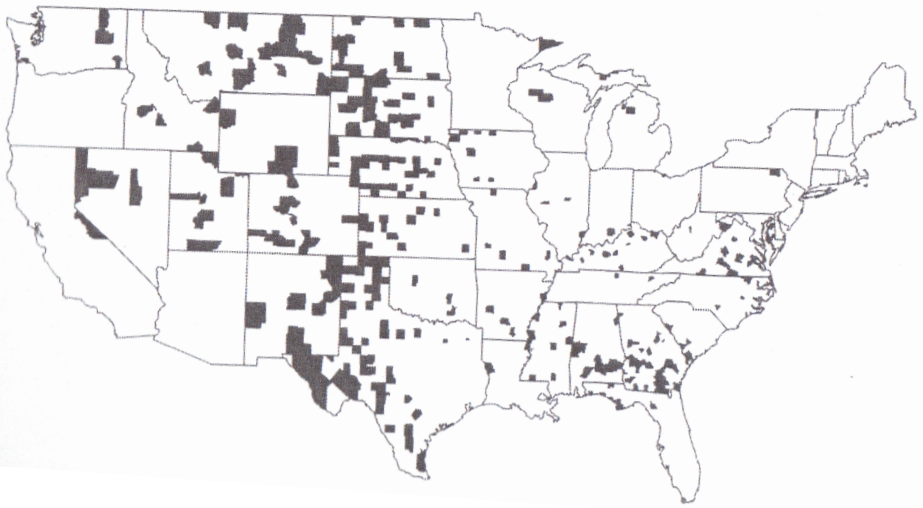


The counties of the US with the lowest 10% age-standardized death rates for cancer of kidney of US white males, 1980-1989.

Highest kidney cancer death rates



Lowest kidney cancer death rates



Methods for Statistical Analysis

- What is the norm in the field?
- A spectrum of alternative statistical methods

Bias, Inappropriate method

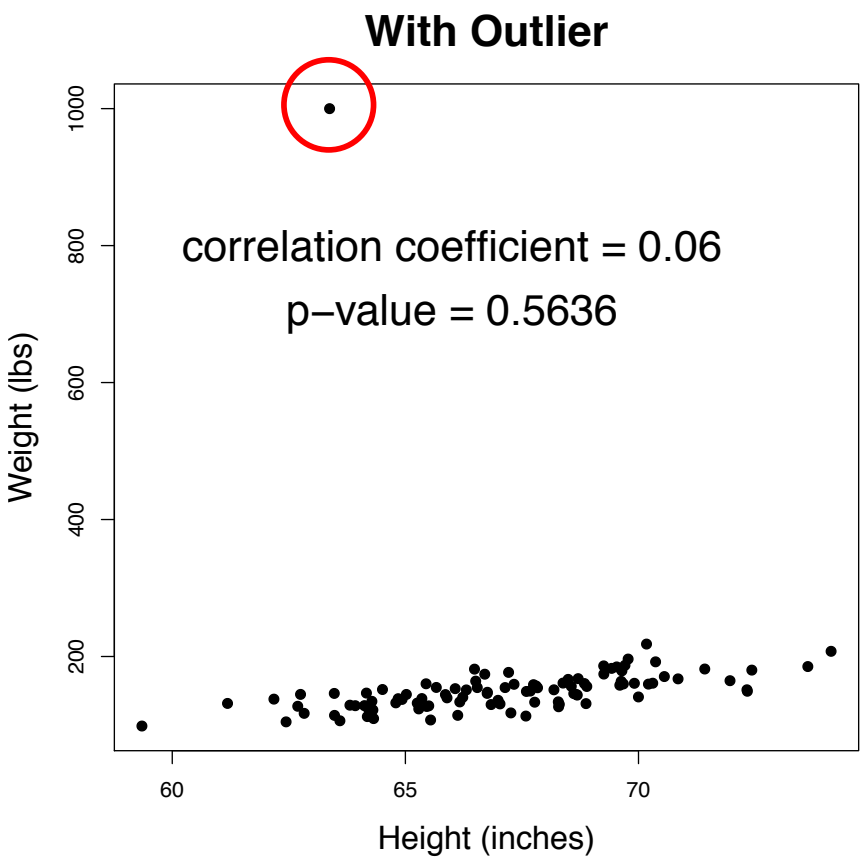
General method with stated assumptions

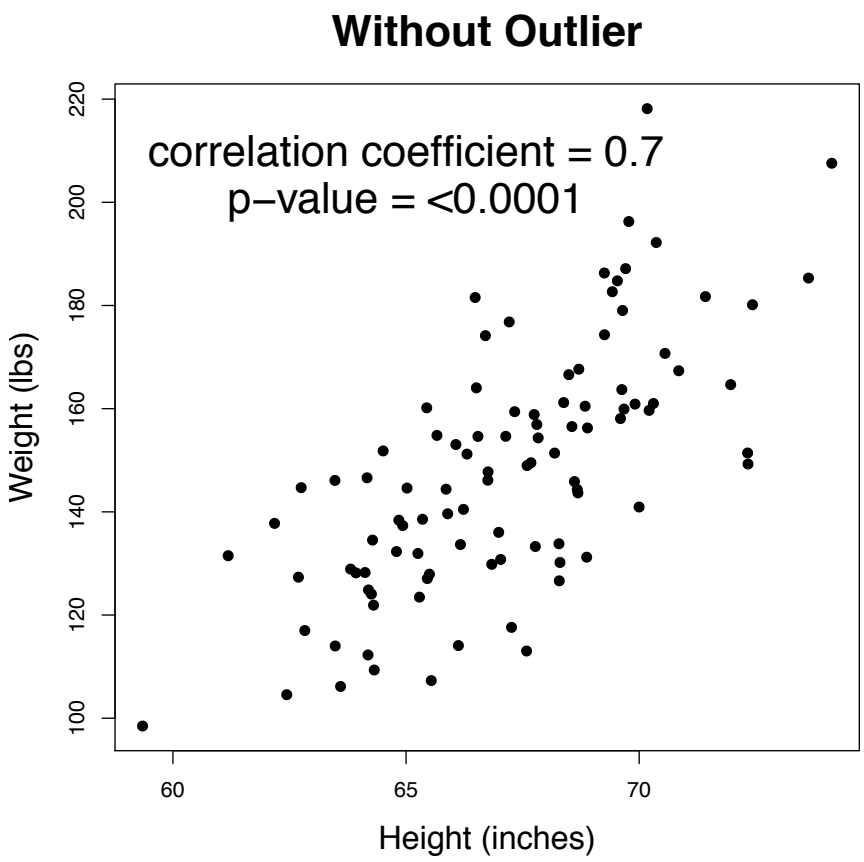
Most statistically rigorous method that evaluates most/all assumptions

Increasing scope
Increasing monetary and time costs
Increasing precision

GRAPHING YOUR DATA

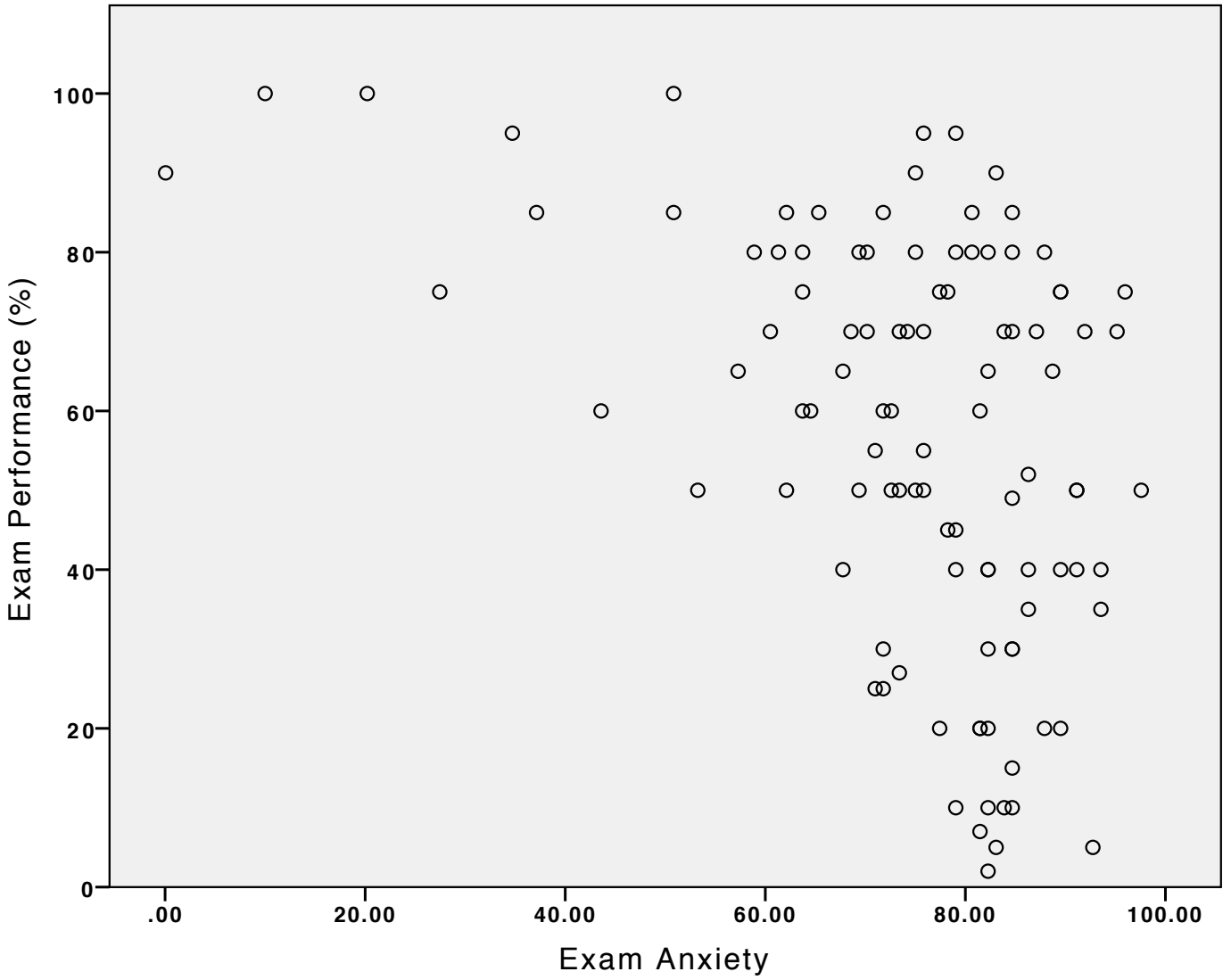
Before Analysis – e.g., Outliers



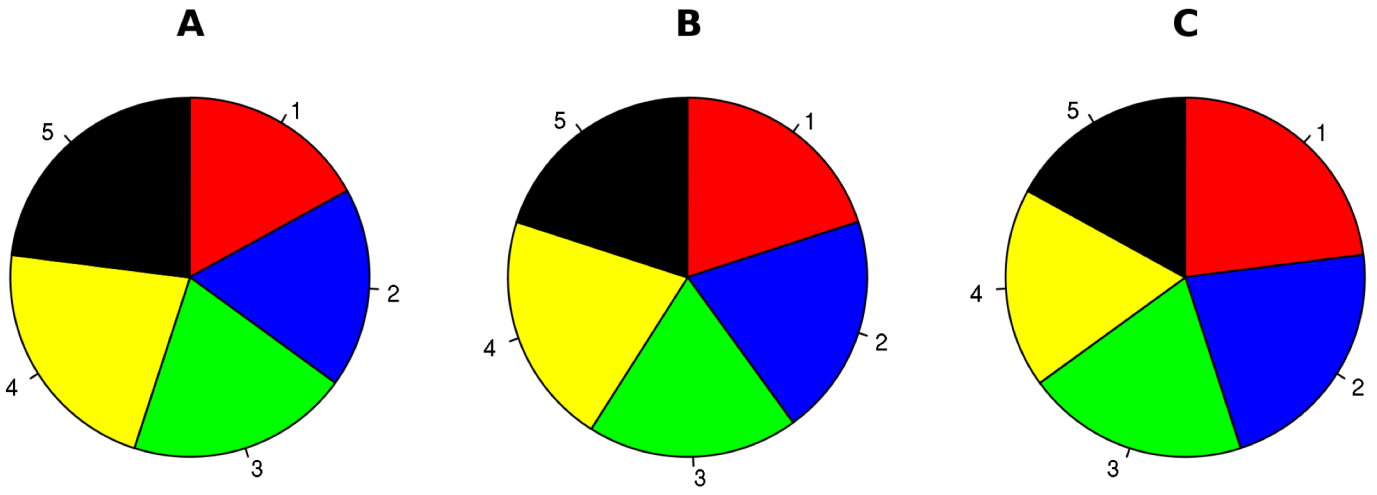


Before Analysis

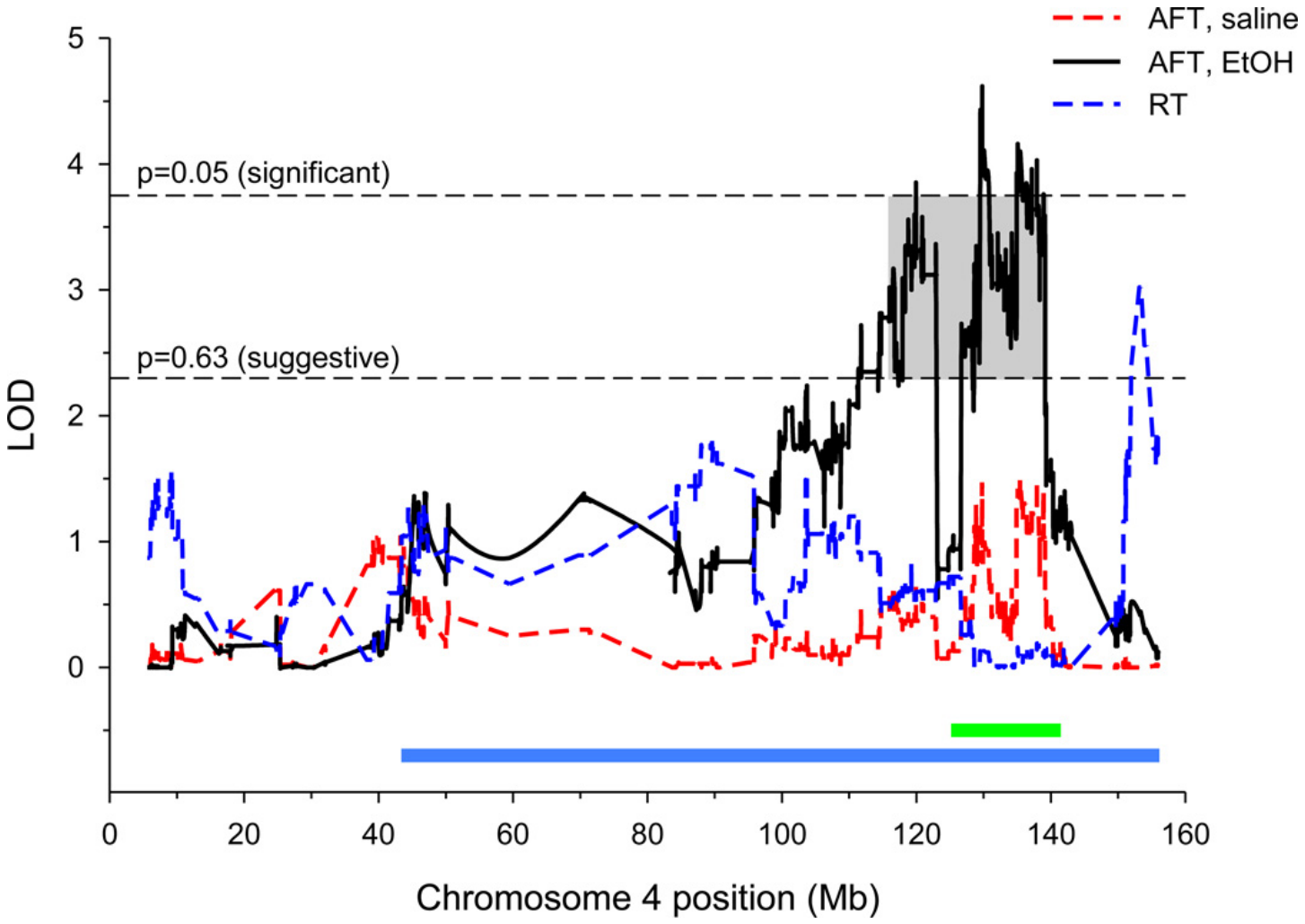
understanding your data



After Analysis



After Analysis



Bennett B, Larson C, Richmond PA, Odell AT, Saba LM, Tabakoff B, Dowell R, Radcliffe RA. Quantitative trait locus mapping of acute functional tolerance in the LXS recombinant inbred strains. Alcohol Clin Exp Res. 2015 Apr;39(4):611-20.

WHERE TO SEEK HELP

Your Resources

- The book
- Google
- Other students/lab mates
- Me
- Colorado Biostatistics Consortium