

Data Visualization Strategies

Justin Joque
Nicole Scholtz
Clark Data Labs 2015

Overview

- Introduction
- Types of visualizations
 - Example I
- Thinking about complex data
- Dealing with complex data
 - Example II
- More types of visualizations
 - Example III
- Problems and exotic solutions
- Conclusions

Introduction

Starting out

- What are you trying to show?
- What relationships are you interested in?
- What information is unnecessary?
- What do you know about your data?
- Who is your audience?
- What have other people done?

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray et le journal intérieur de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Malibow se sont rejoints vers Orscha et Witebsk, avant de toujours marcher avec l'armée.

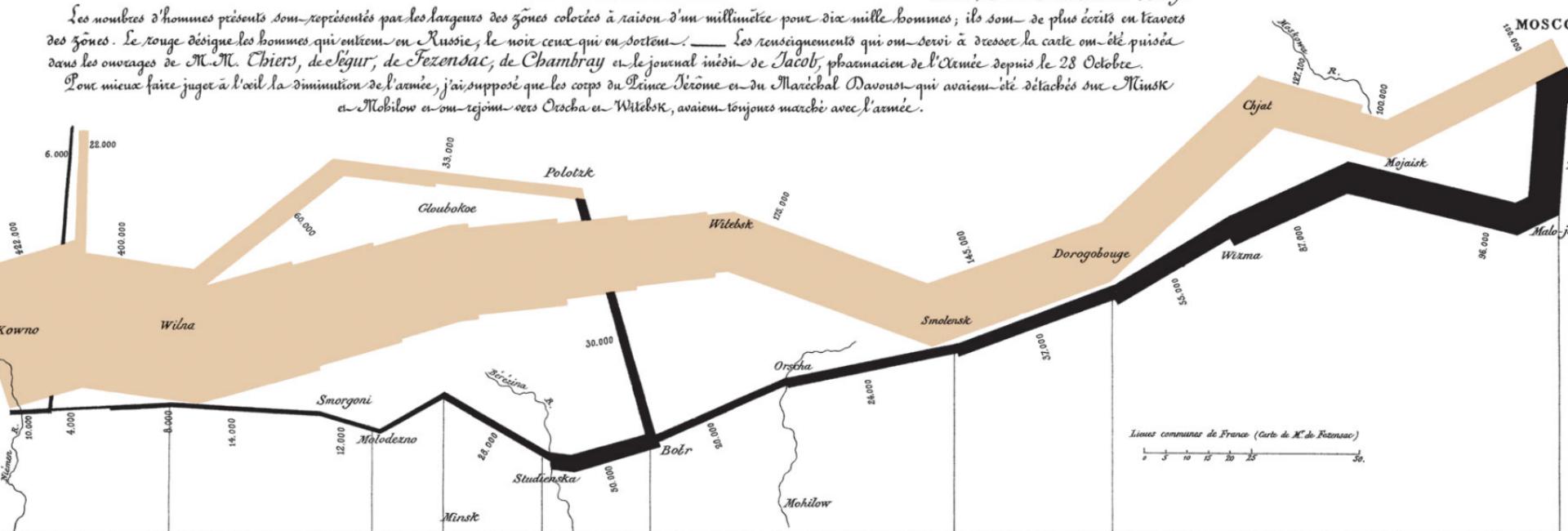
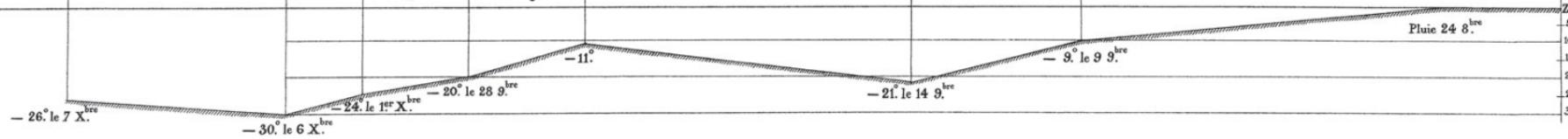
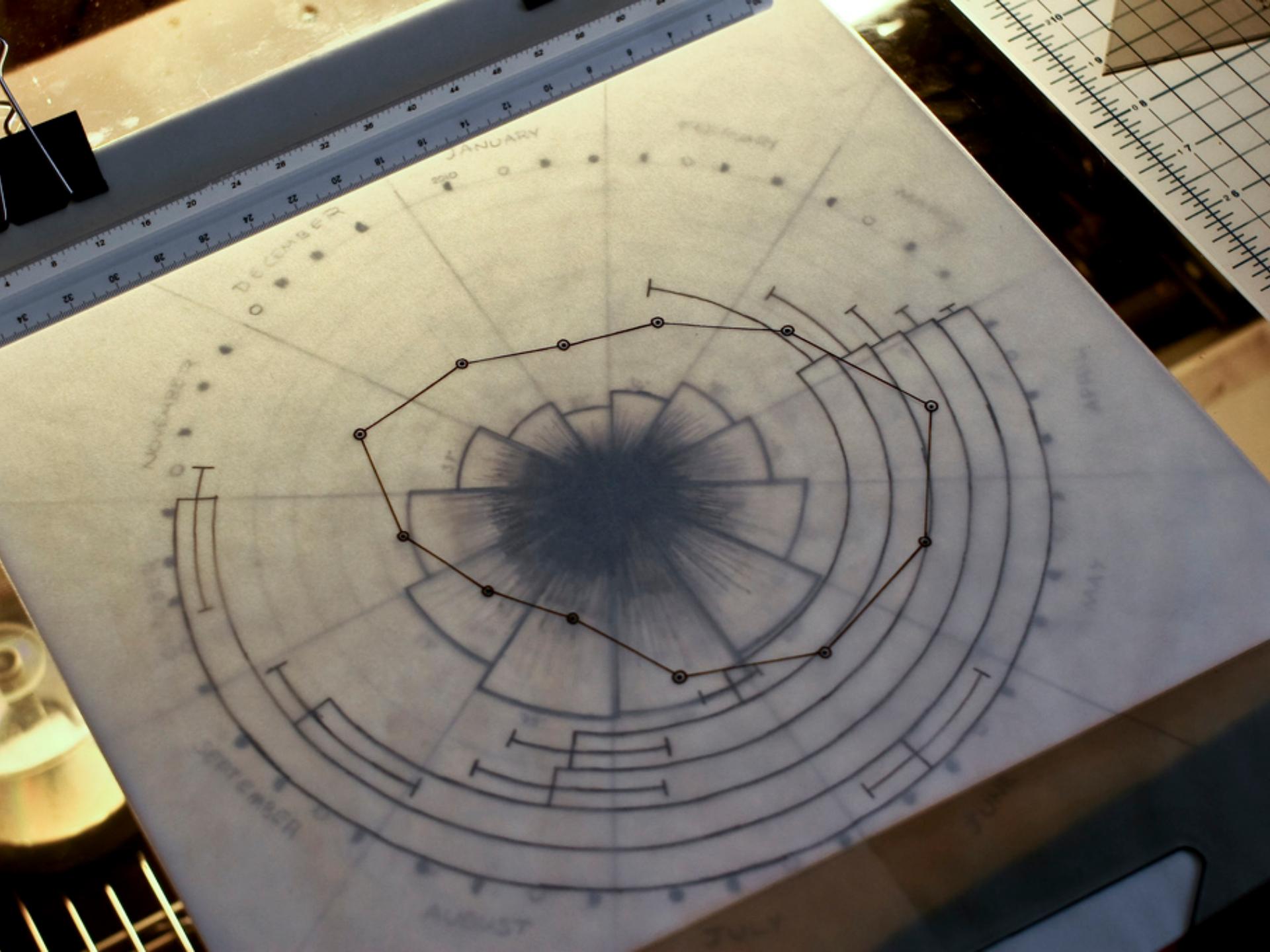
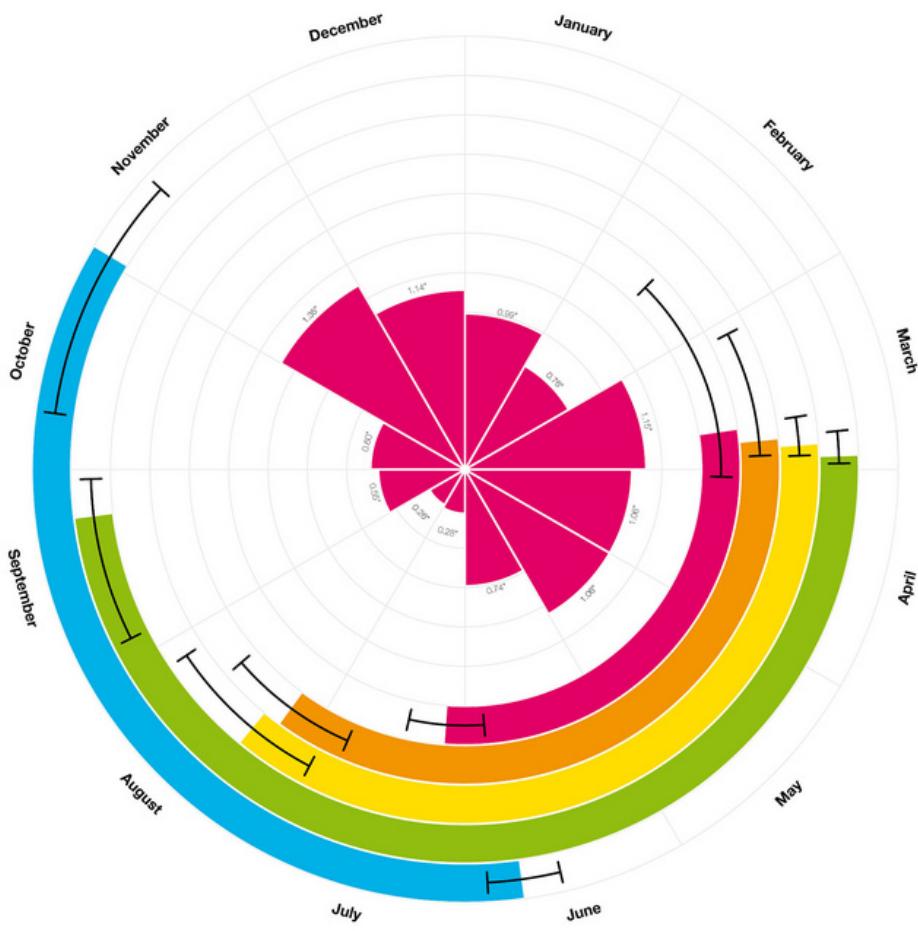


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

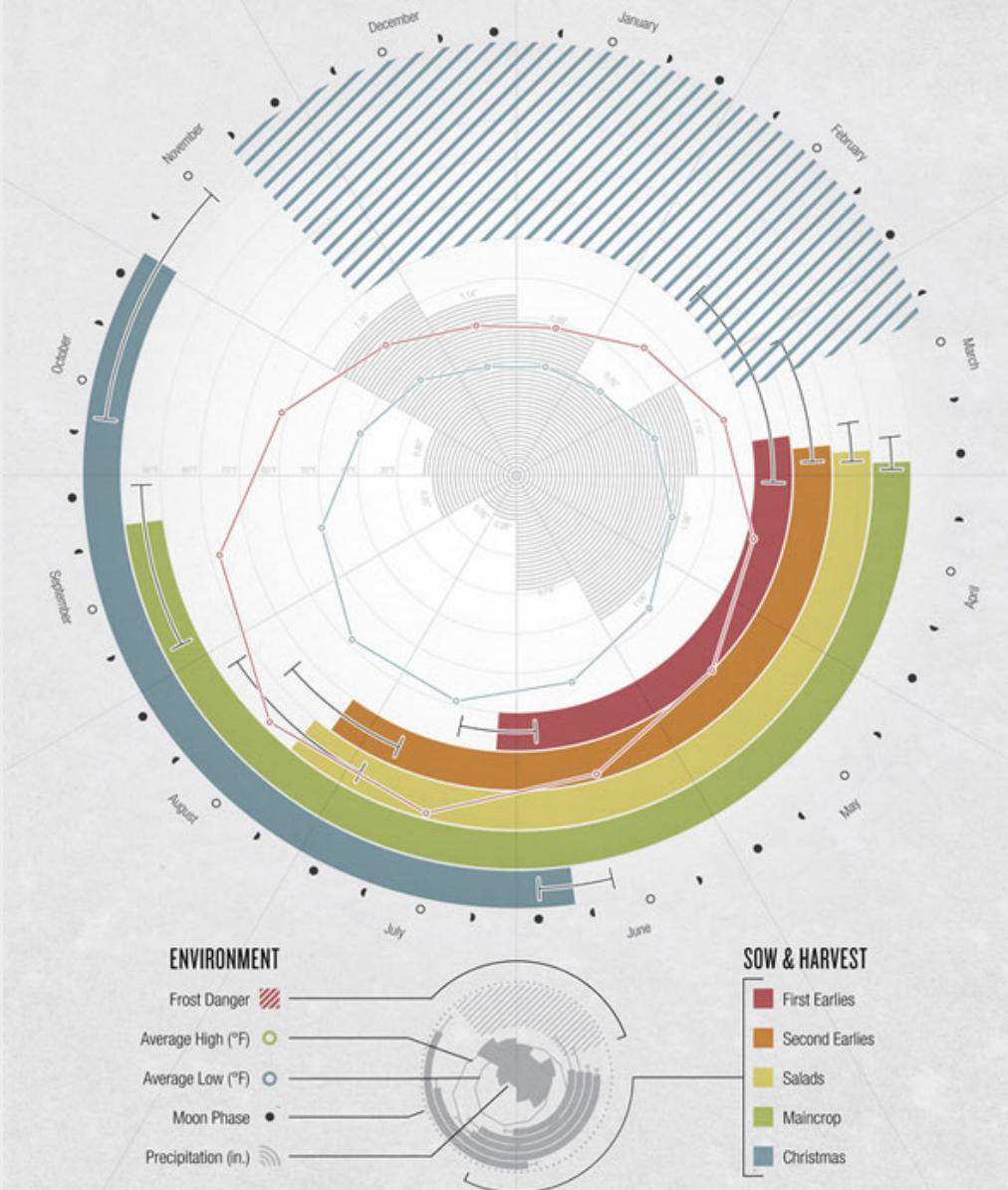






THE POTATO LIFECYCLE

A Comprehensive Timetable For Planning, Planting And Monitoring Your Own Crop Of The World's Most Resilient Tuber



Aesthetics



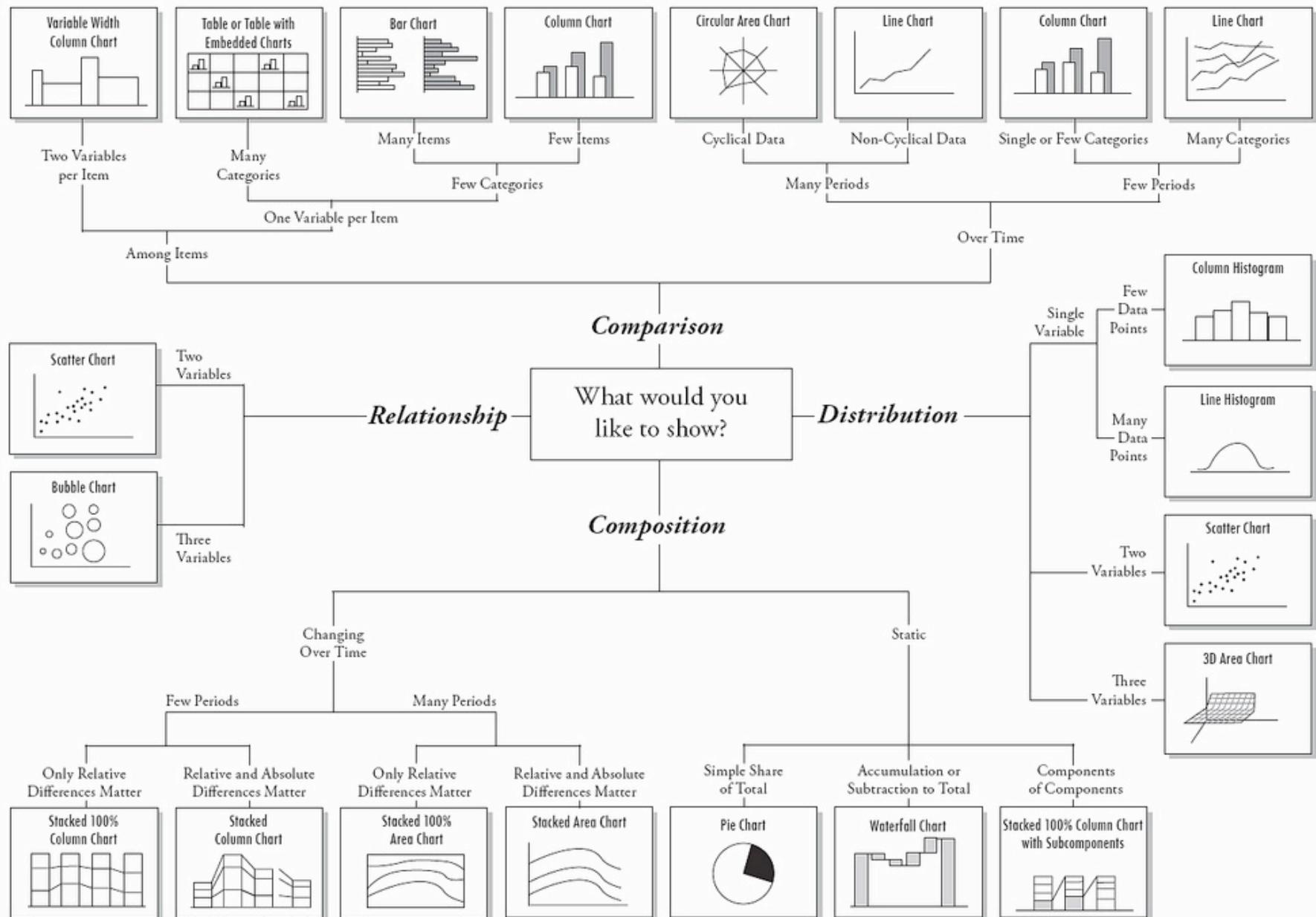
Resolution 5000 x3500 px - free download - www.psdgraphics.com

Less is More

Approaching Visualization

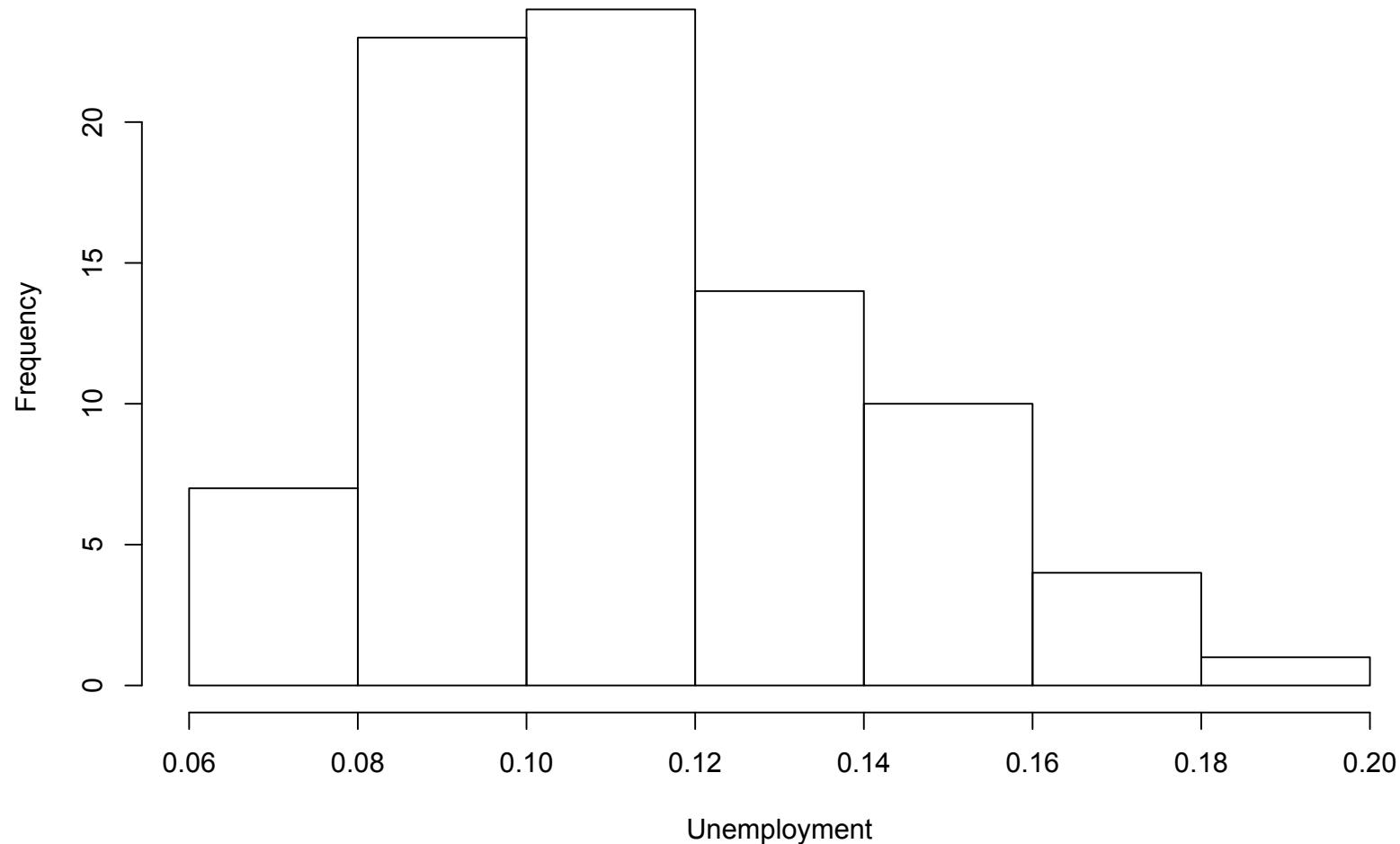
- Organize your data for your analysis or visualization
- Design your visualization around your research question
- If your visual is boring, get better data

Types of Visualizations



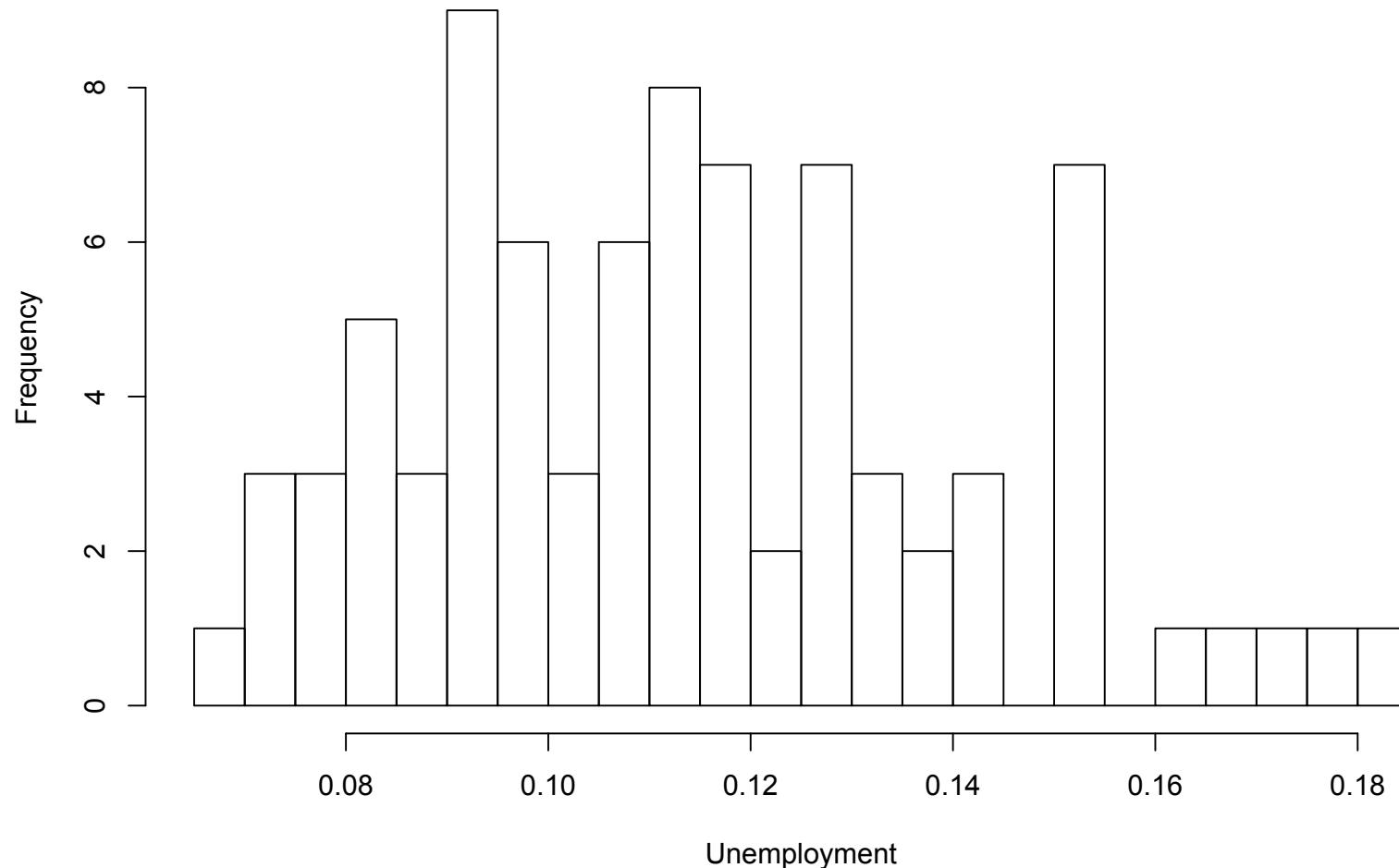
Histogram

Histogram of Unemployment by County; 6 Breaks



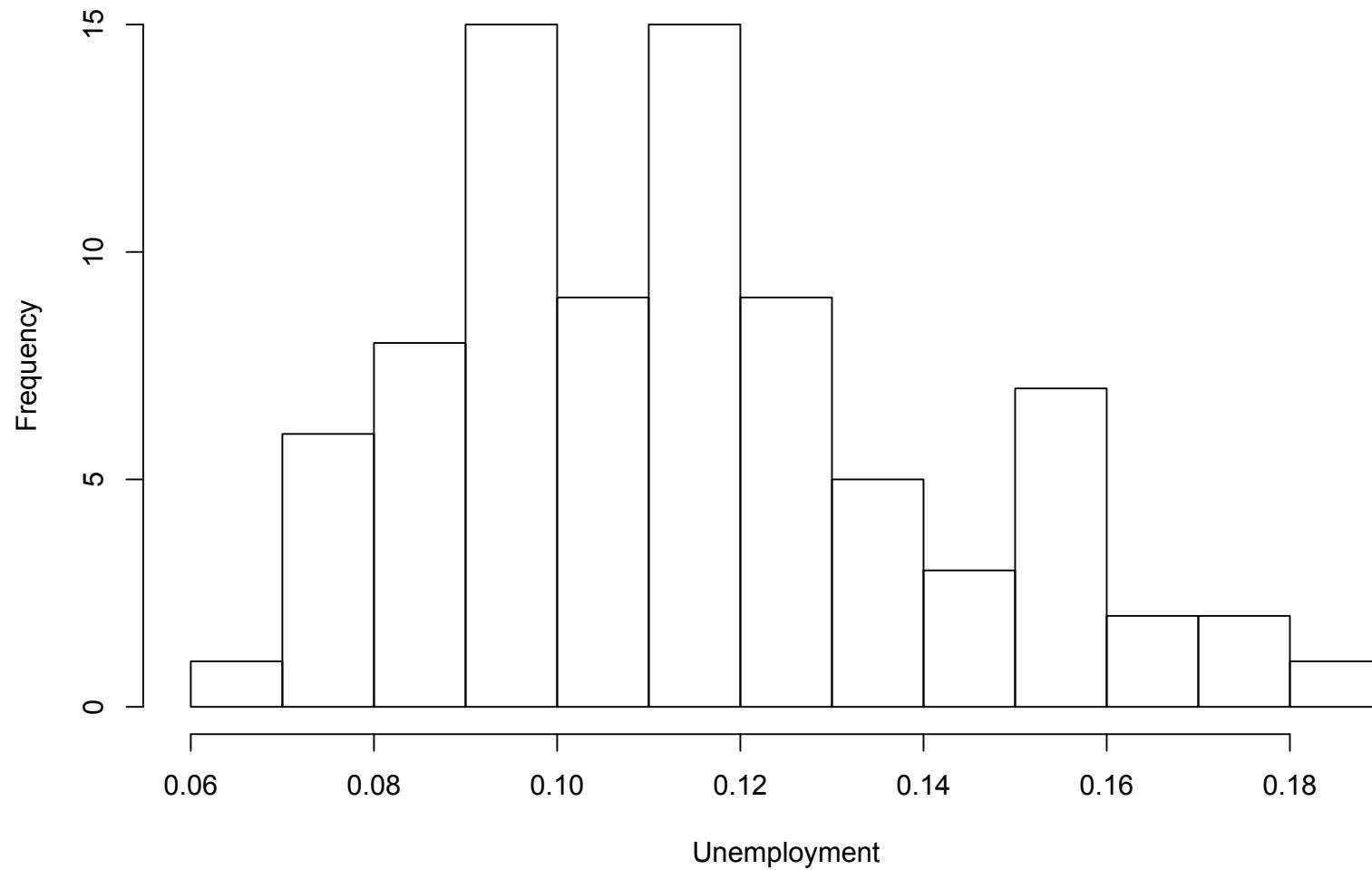
Histogram

Histogram of Unemployment by County; 20 Breaks



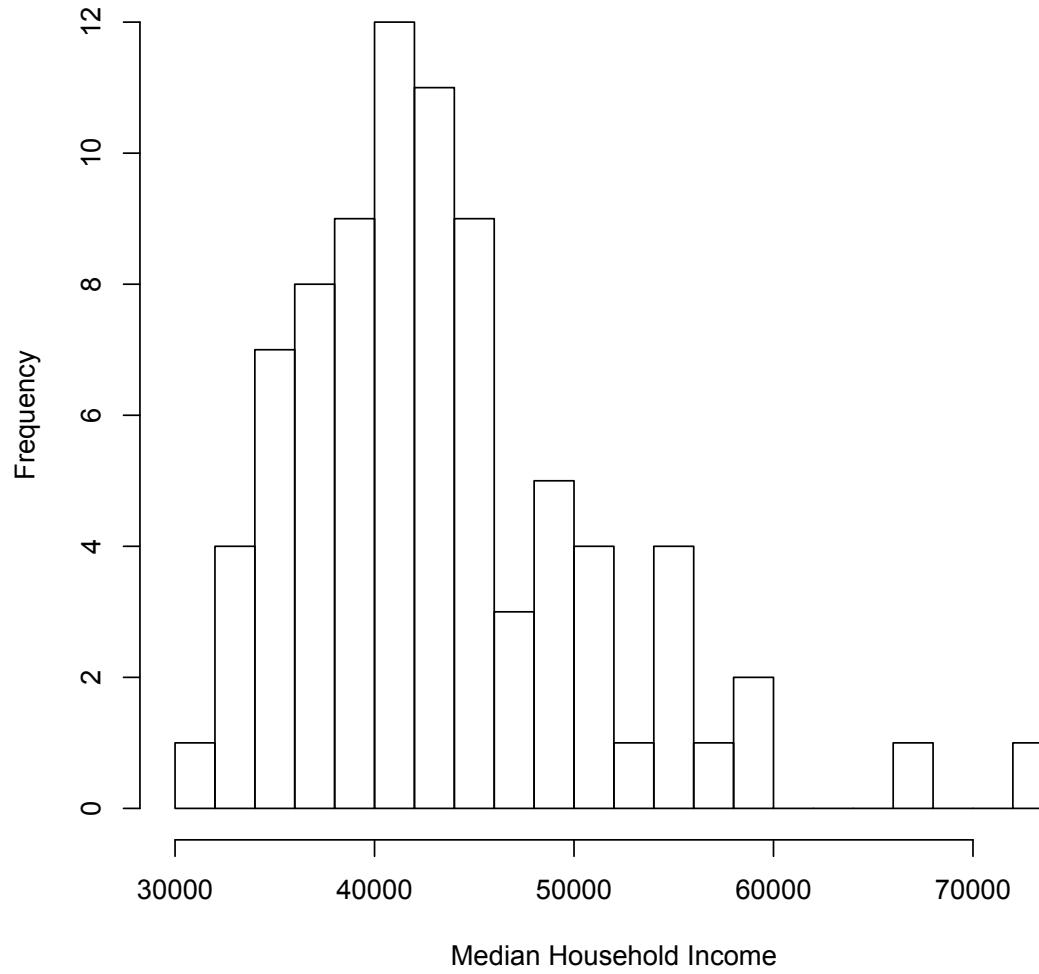
Histogram

Histogram of Unemployment by County; 15 Breaks



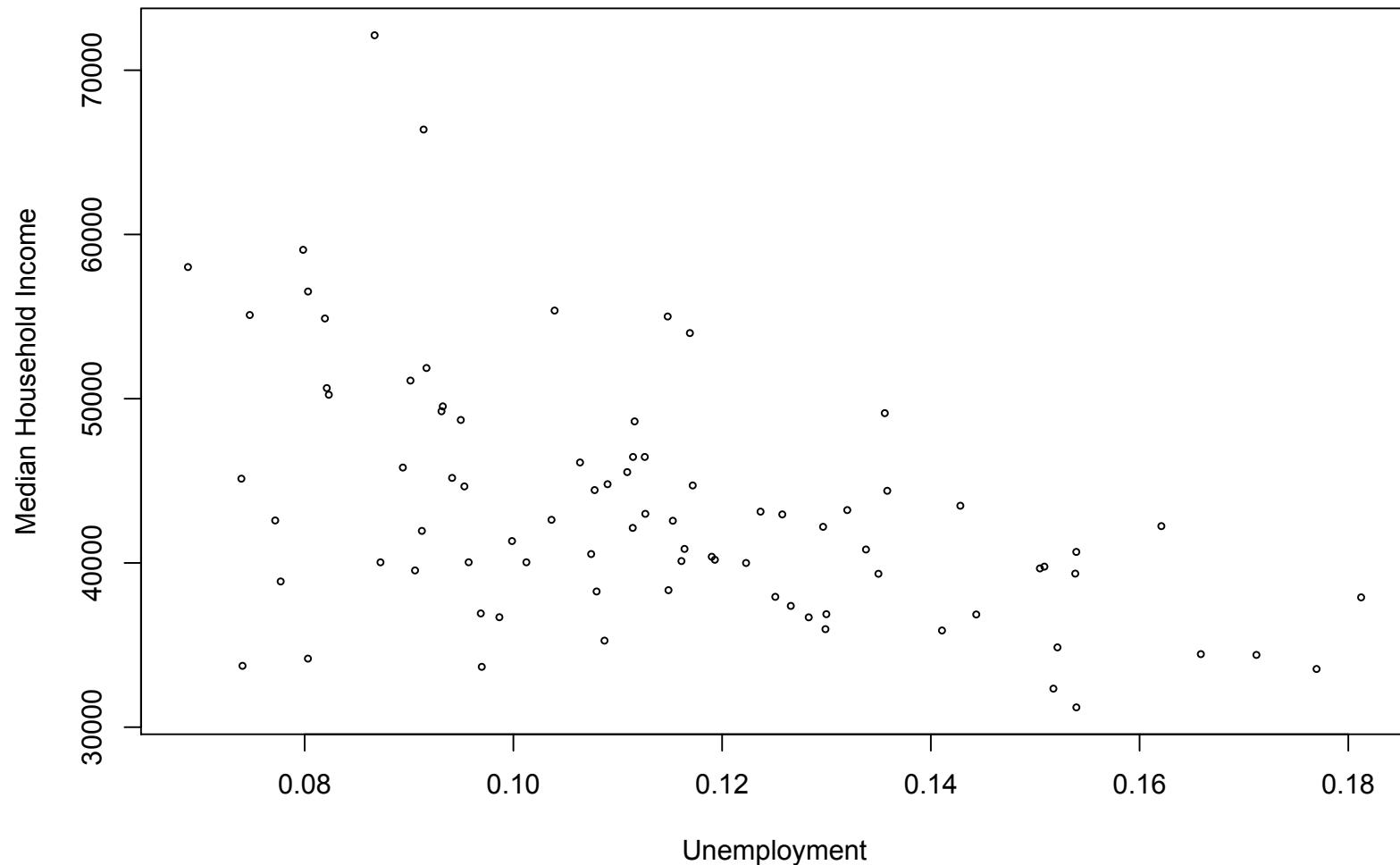
Histogram

Histogram of Median Household Income by County; 20 Breaks



Scatterplot

Median Household Income and Unemployment by County in Michigan, 2010



What is R?

- Free!
- Framework for ‘data programming’
- Can either use scripts or commands
- Uses commands/functions with parameters
- Both = and <- assign data to a variable/data frame
- Large user group has created lots of plug-ins

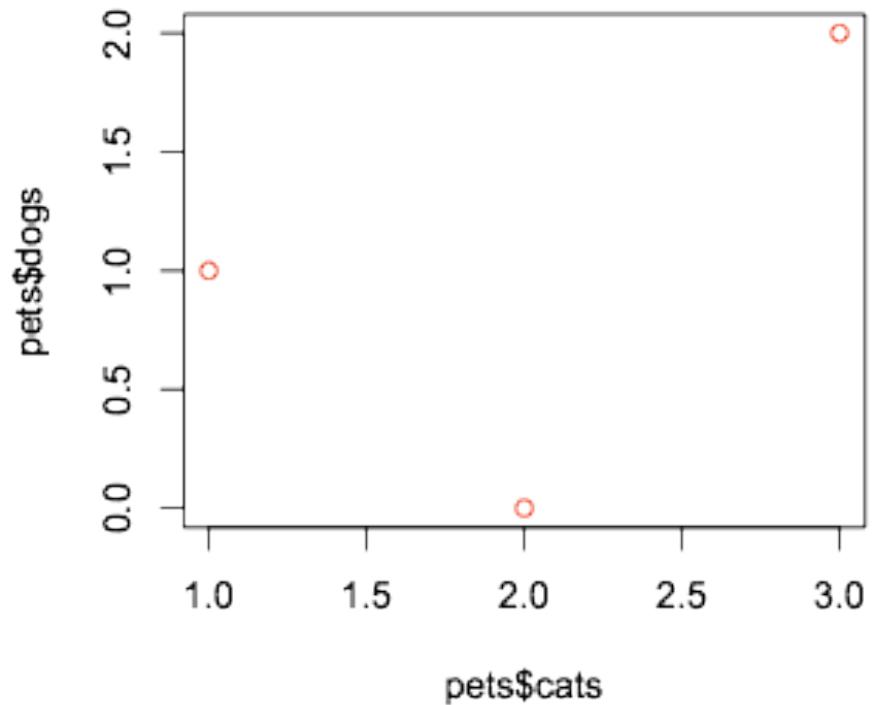
Example I

Use R to create **histograms** and **scatterplots**

Thinking about Data

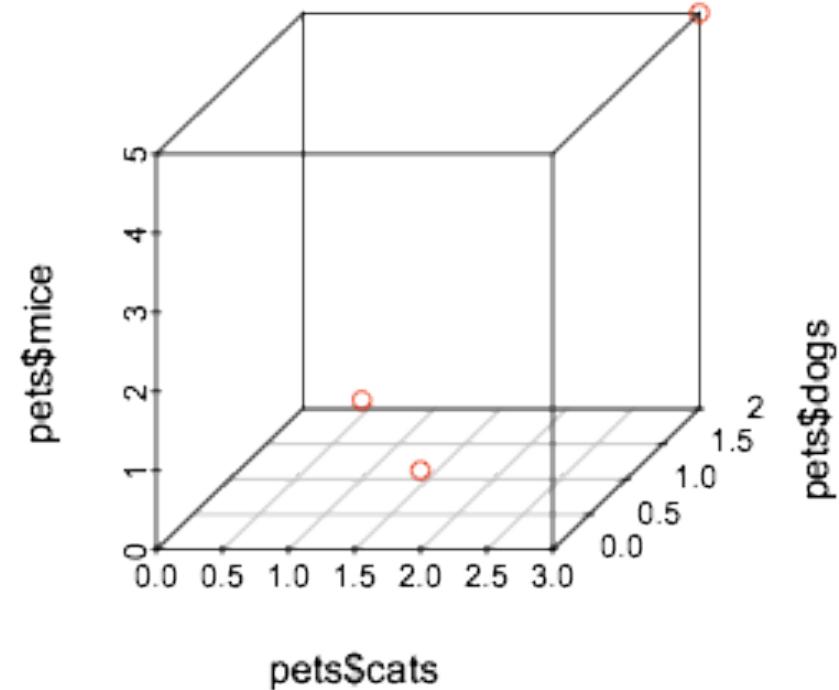
Dimensionality

	cats	dogs
Jane	2	0
John	1	1
Moe	3	2



Dimensionality

	cats	dogs	mice
Jane	2	0	1
John	1	1	1
Moe	3	2	4

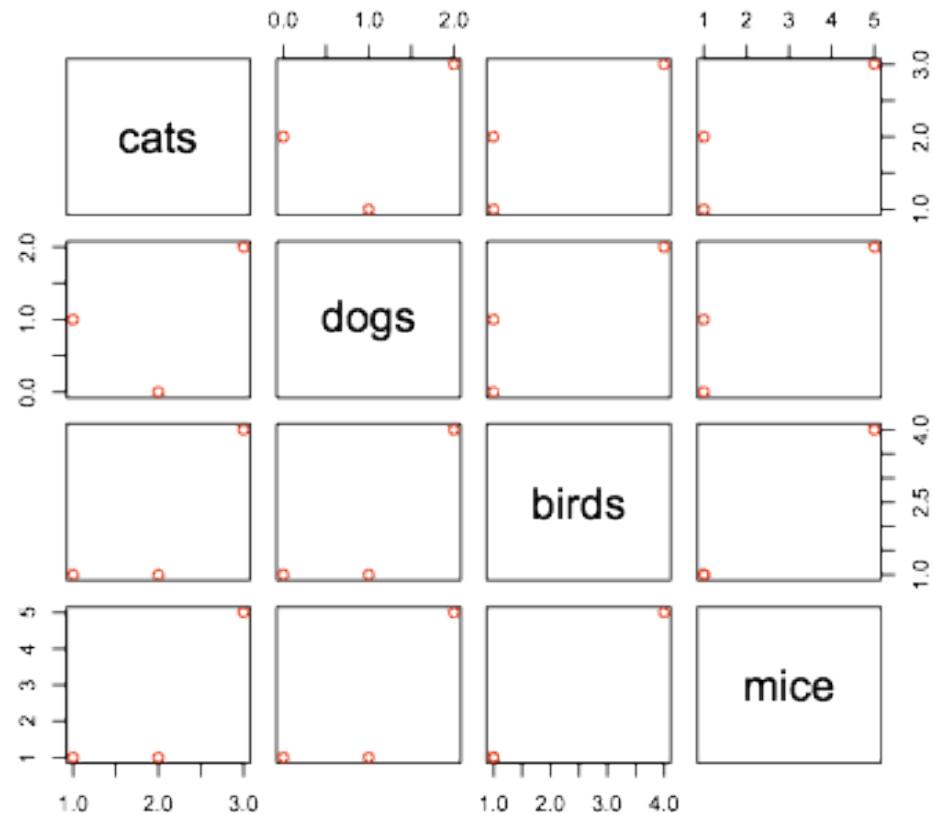


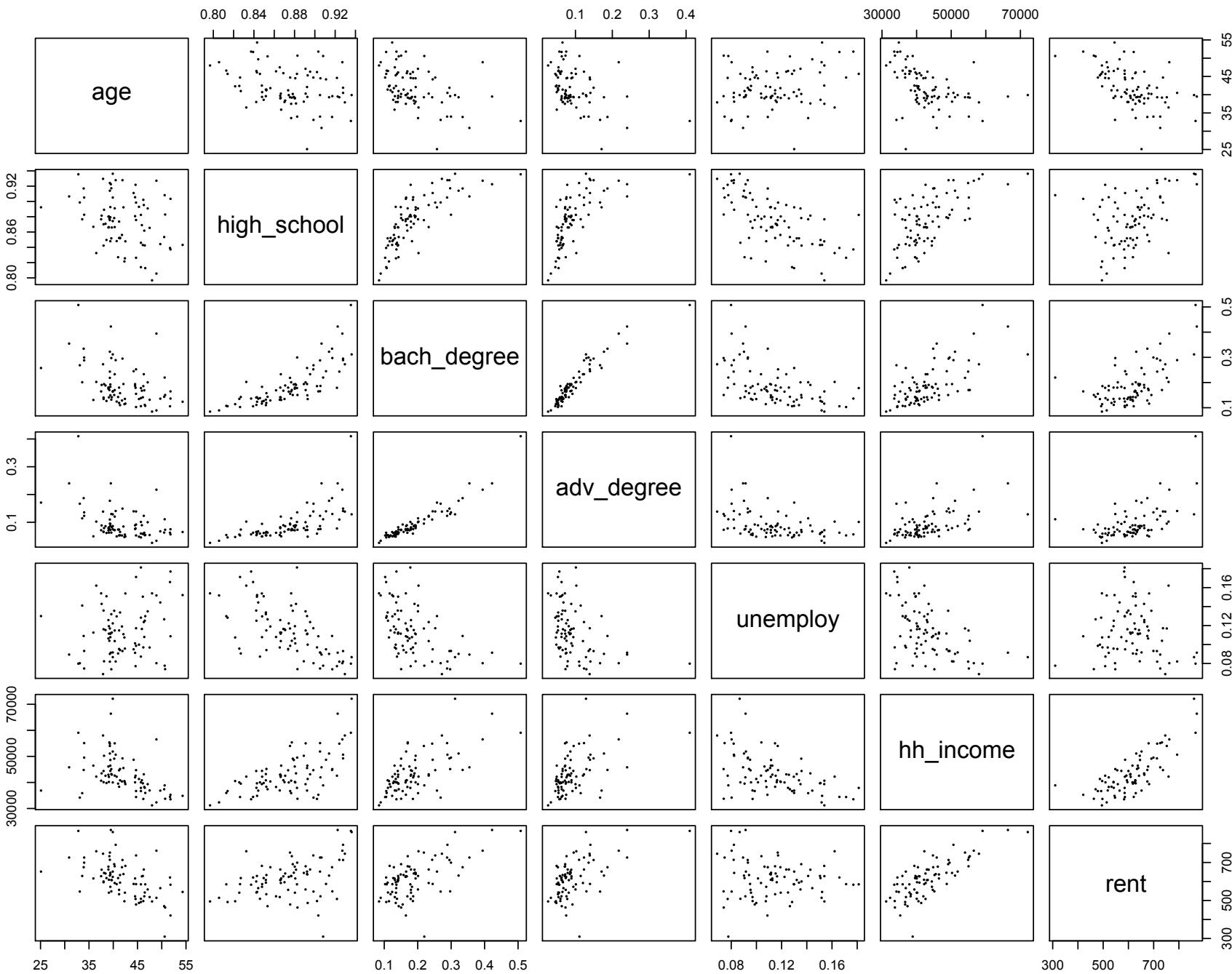
Dimensionality

	cats	dogs	mice	birds
Jane	2	0	1	1
John	1	1	1	1
Moe	3	2	4	5

Dimensionality

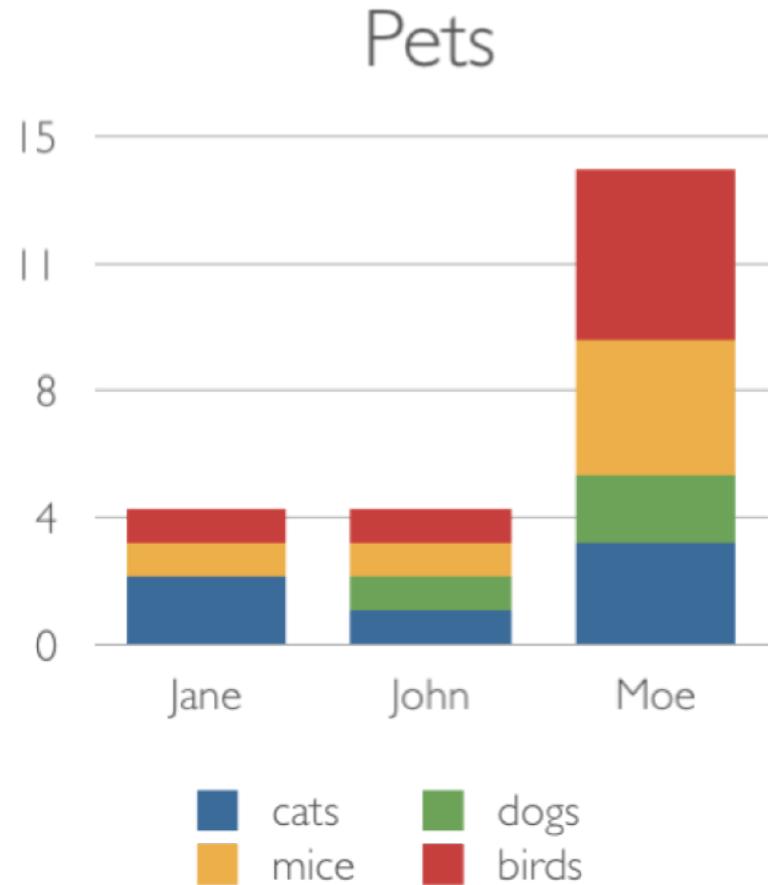
	cats	dogs	mice	birds
Jane	2	0	1	1
John	1	1	1	1
Moe	3	2	4	5





Dimensionality

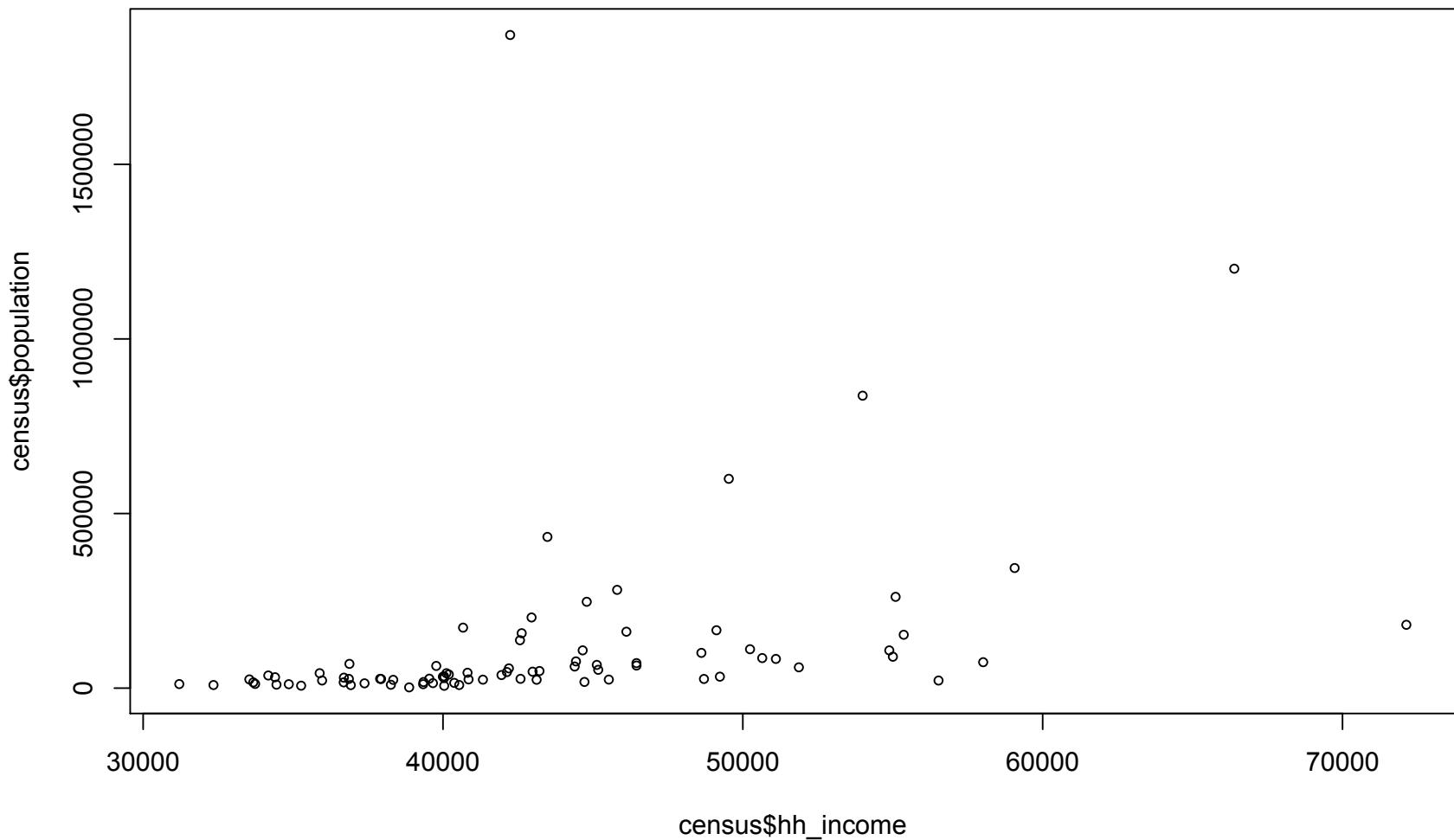
	cats	dogs	mice	birds
Jane	2	0	1	1
John	1	1	1	1
Moe	3	2	4	5



Dealing with Complex Data

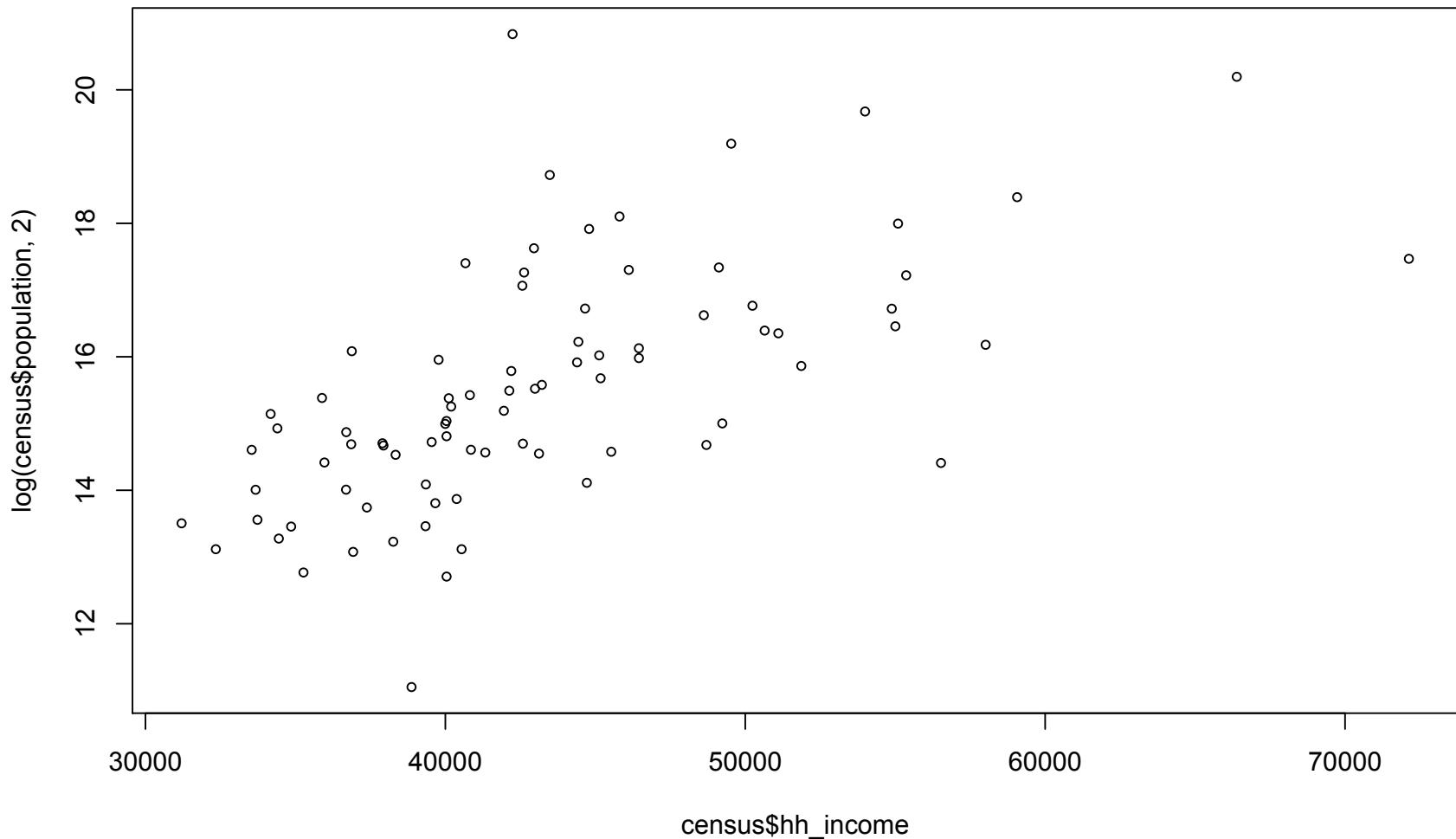
Scaling

Median Household Income and Population by County in Michigan



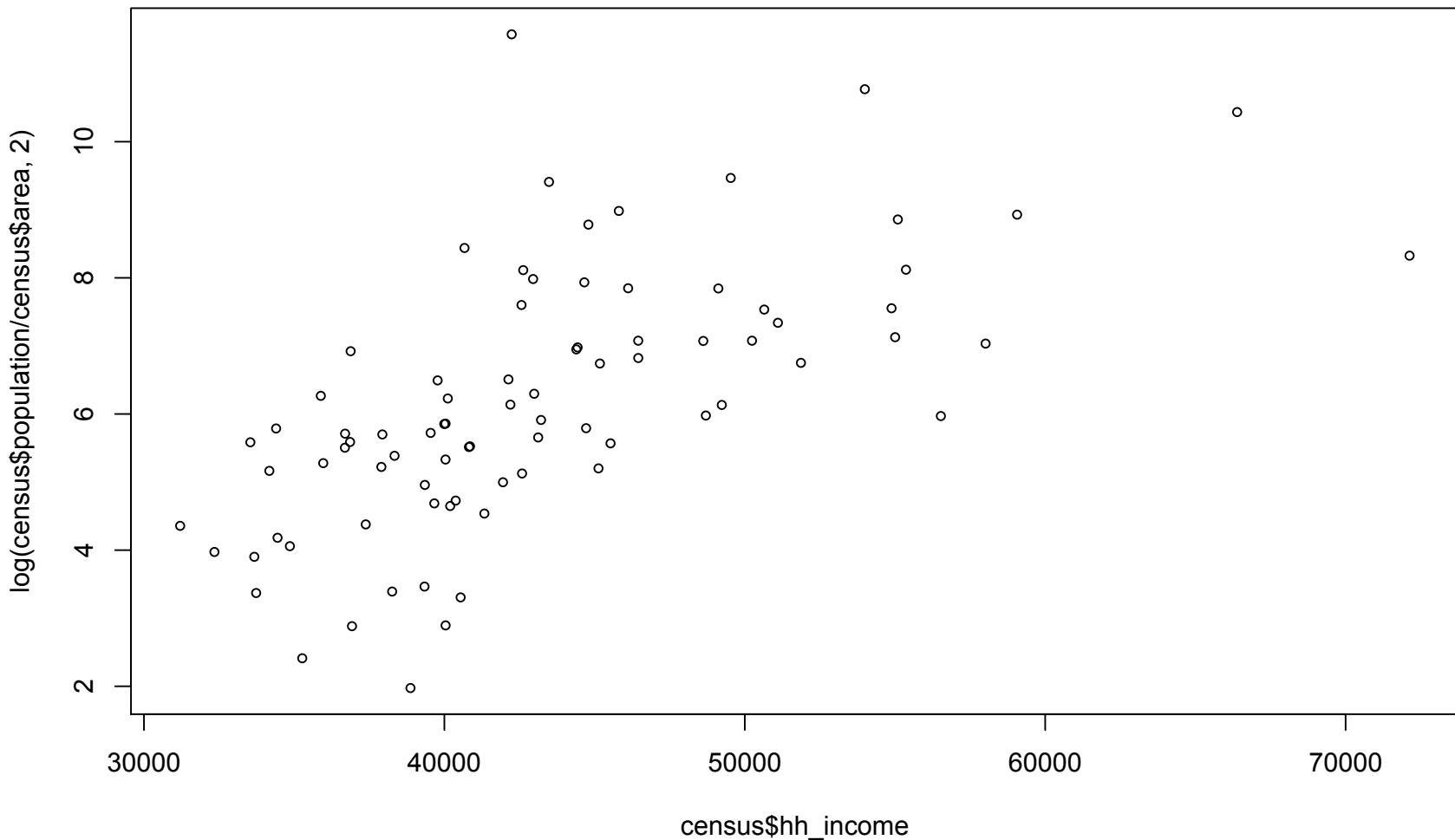
Scaling

Median Household Income and Population by County in Michigan



Normalizing

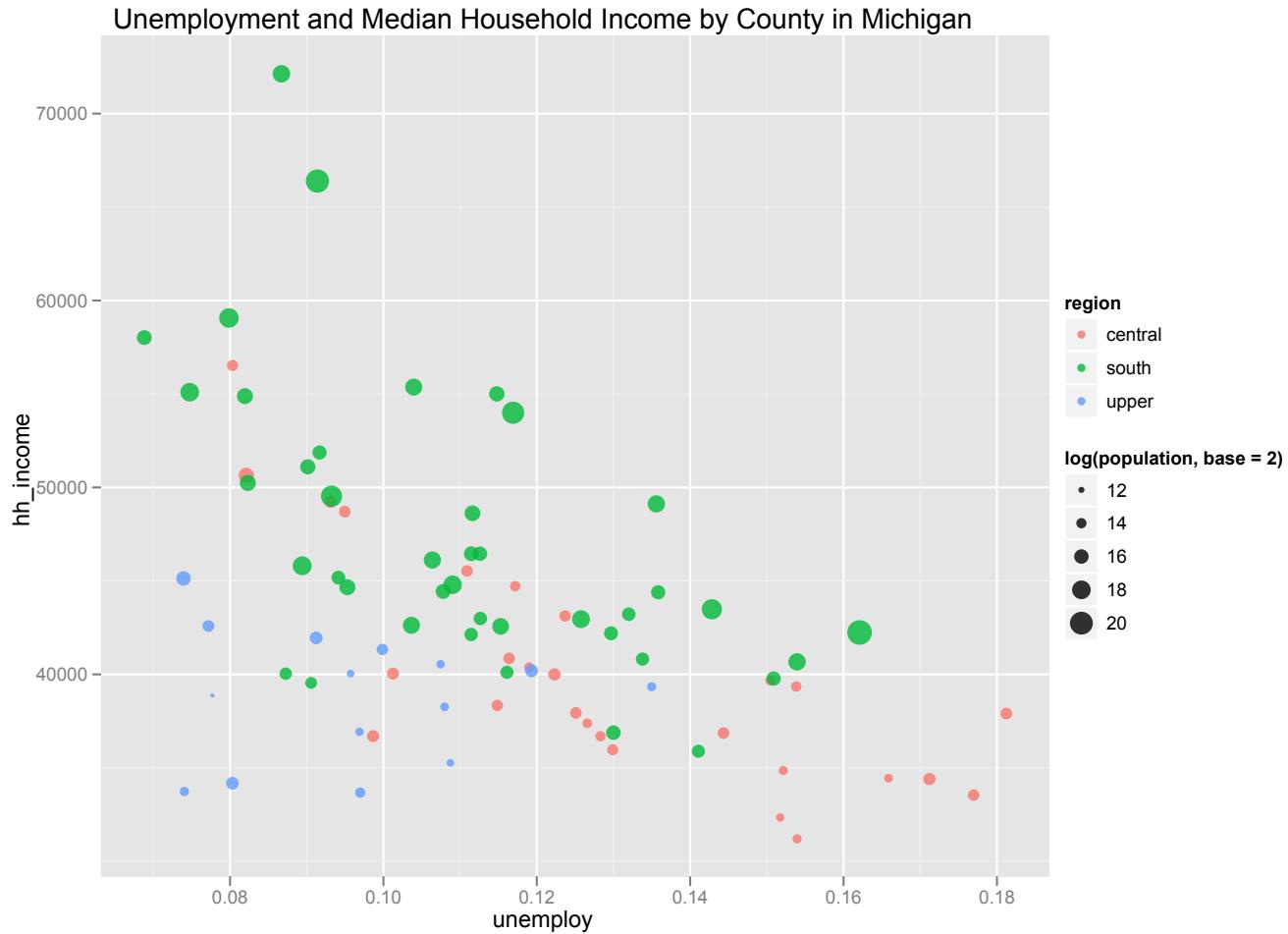
Median Household Income and Population Density by County in Michigan



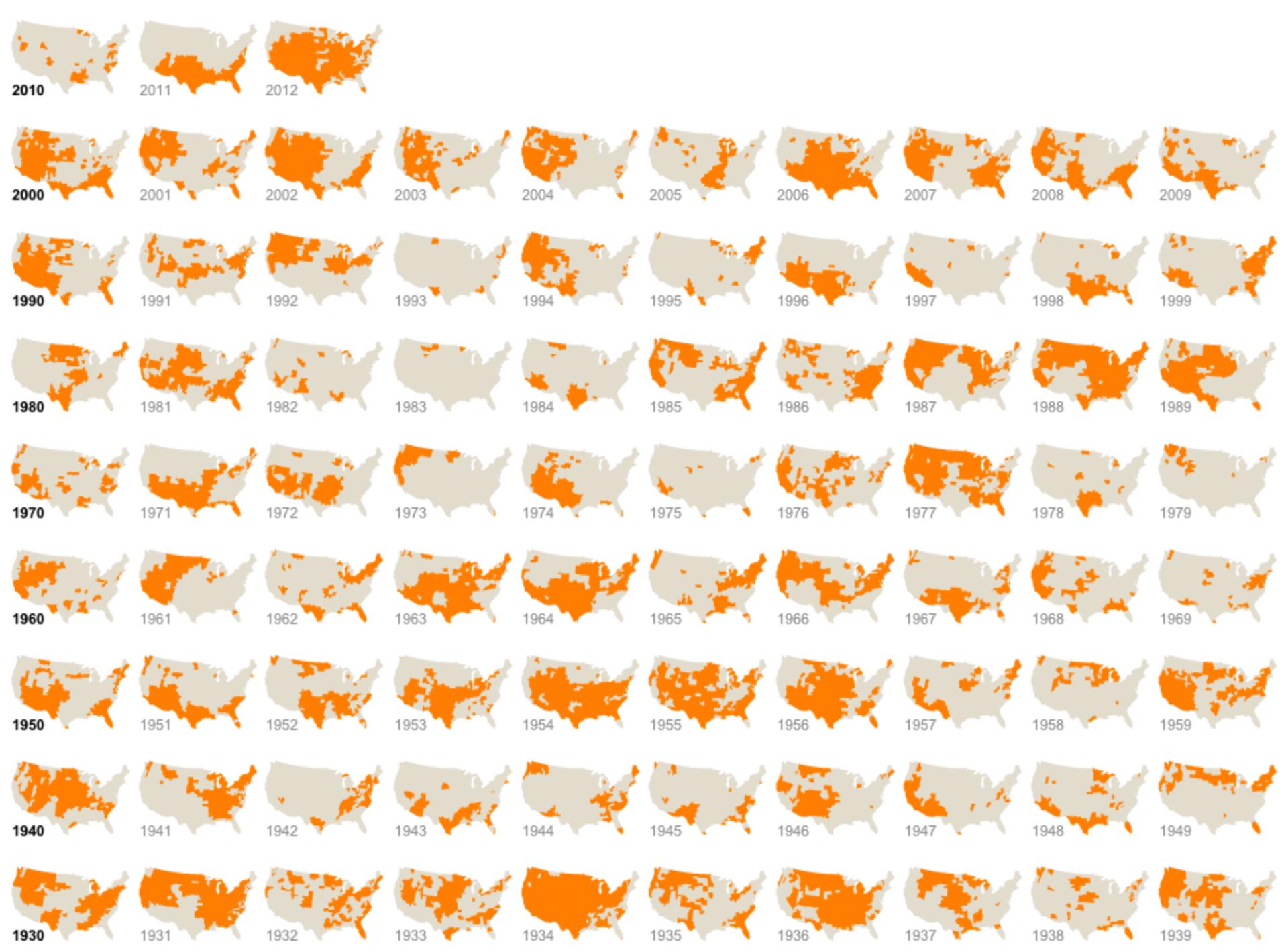
Categorical vs Quantitative

Dimensions

- Color
- Size
- Time

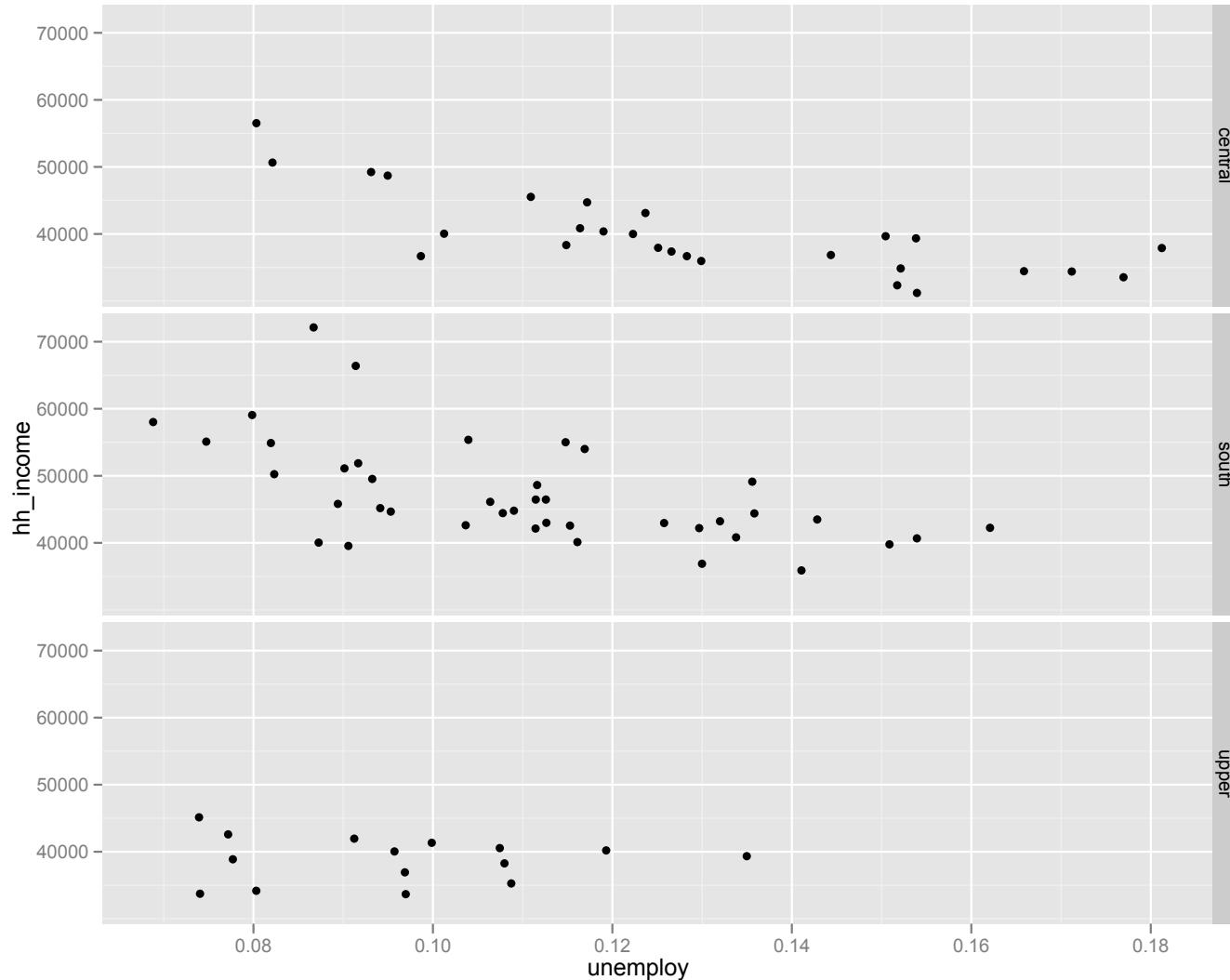


For instance: Gap Minder www.gapminder.org



Grouping Data

Unemployment and Median Household Income by County in Michigan
ACS 2006-2010

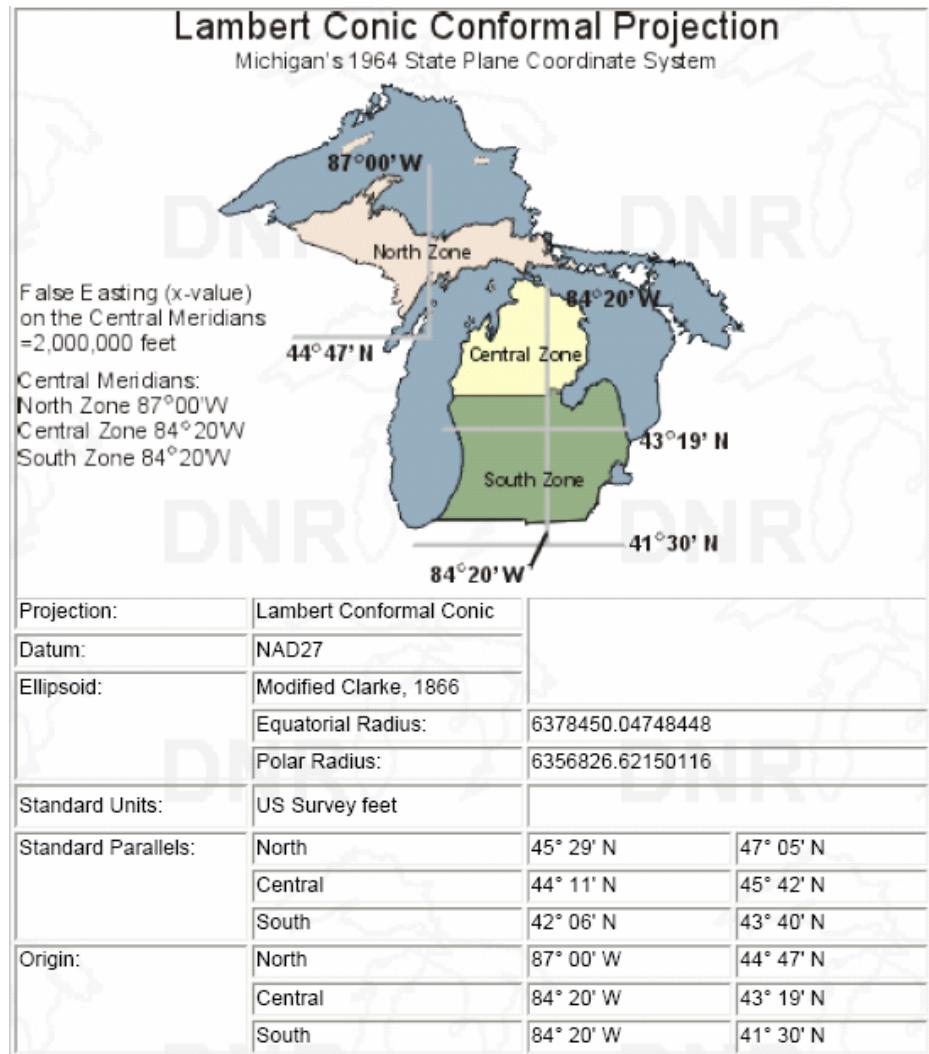


Aggregating data

- Excel pivot tables are great for this
- Use aggregate functions
 - Sum
 - Mean/Median
 - Variance/Standard Deviation
- Often this has been done already (e.g. Census data)

Example II

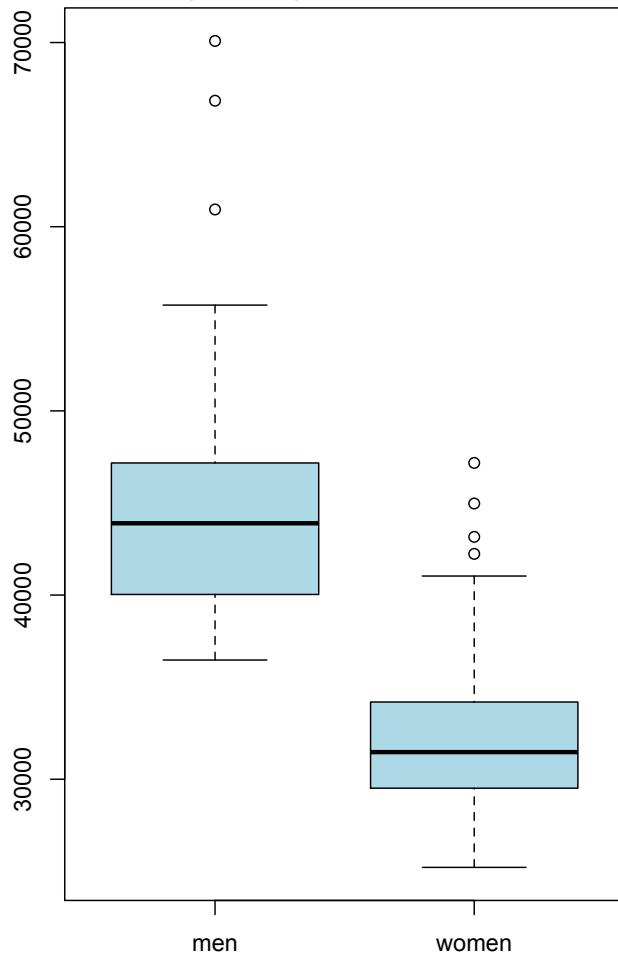
Use the provided data
to make **scatterplots &**
bubble plots



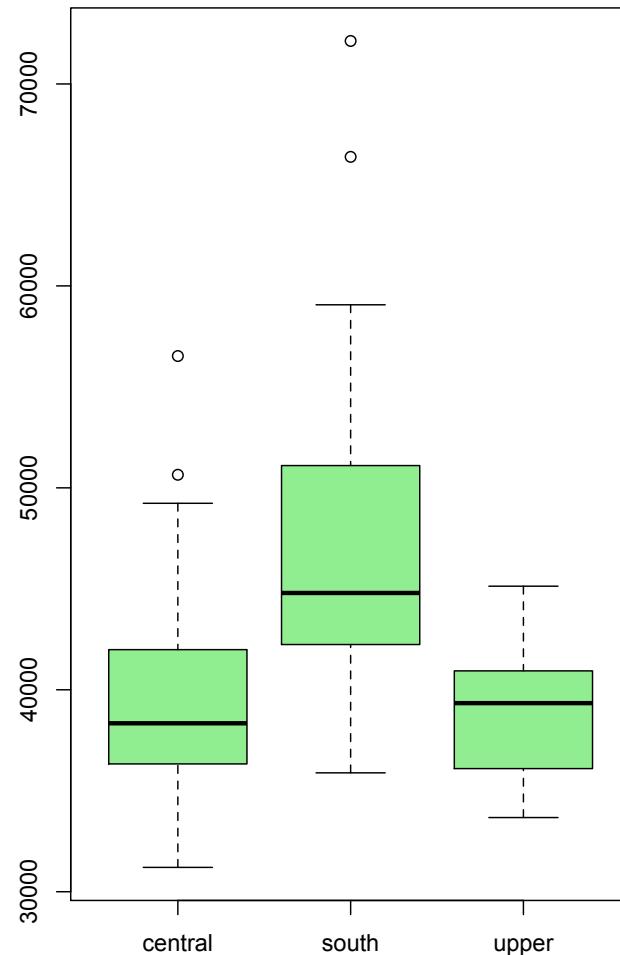
More Types of Visualizations

Box & Whisker Plot

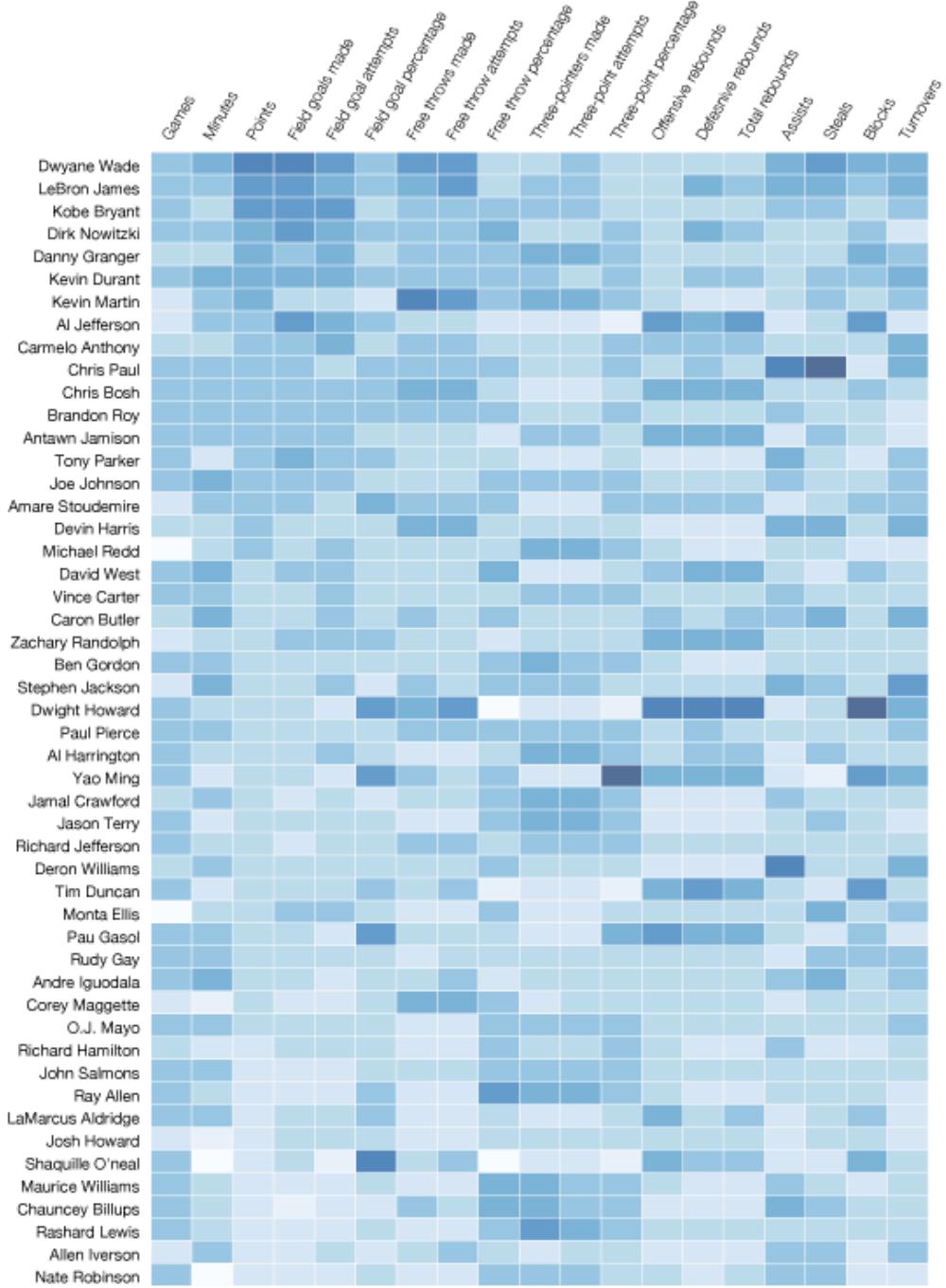
**Median Income by Sex For Population
That Worked Full-time Last 12 Months
by County ACS 2006-2010**



**Median Income by Region
by County ACS 2006-2010**



Heatmap



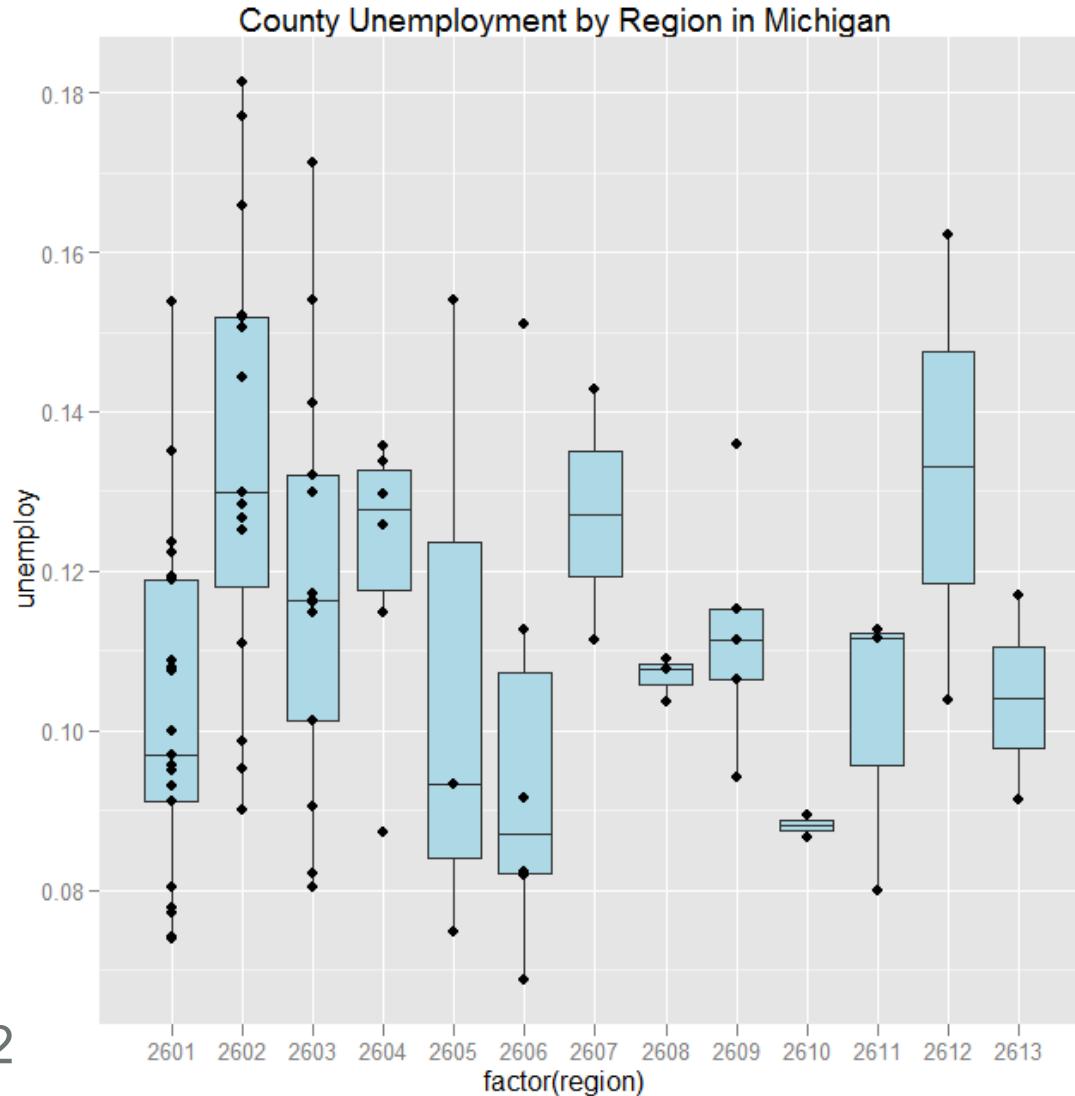
Source: databaseBasketball

Example III

Use the same data to create
boxplots & heatmaps

Problems & Exotic Solutions

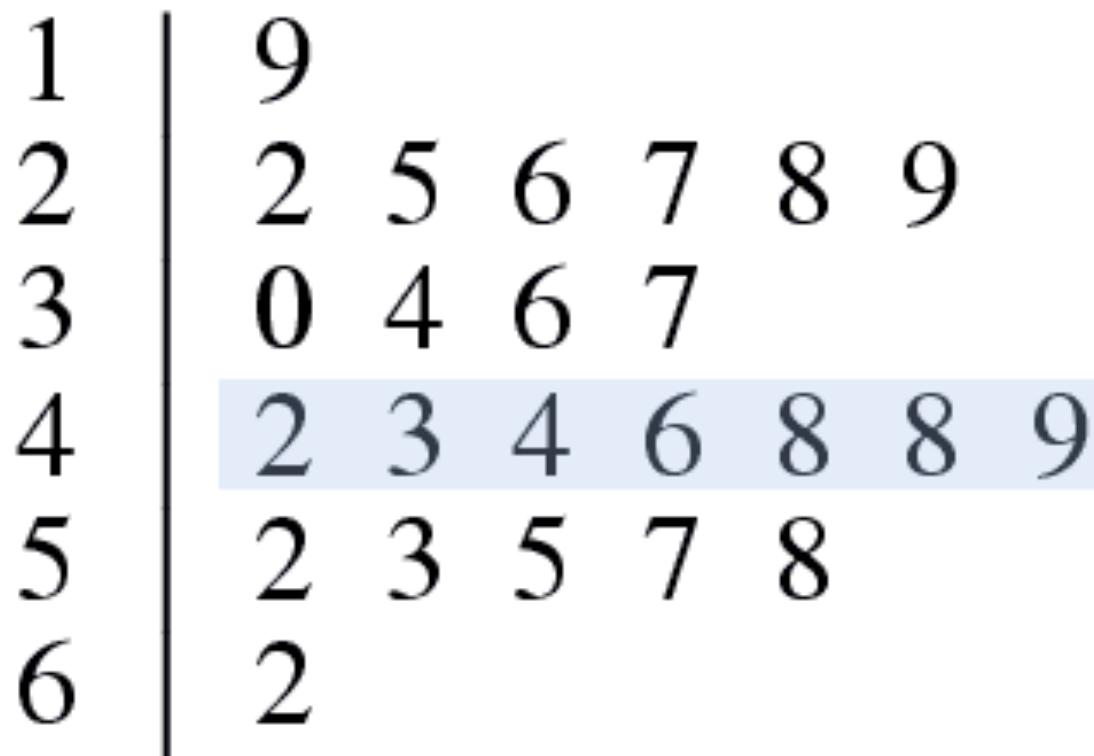
Layered Graphs



Library: ggplot2

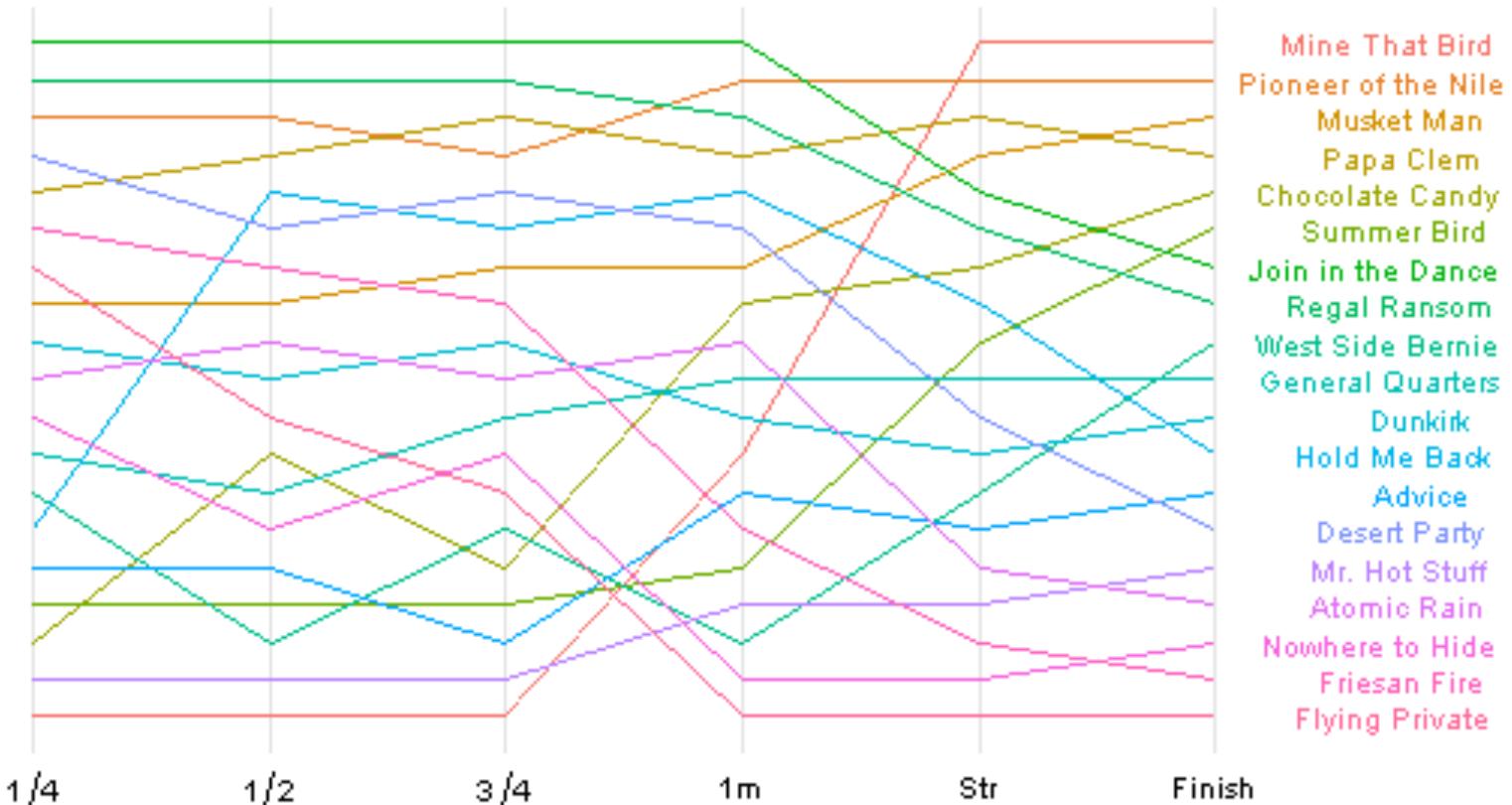
Layered Graphs

Stem and Leaf Plot



Ranking

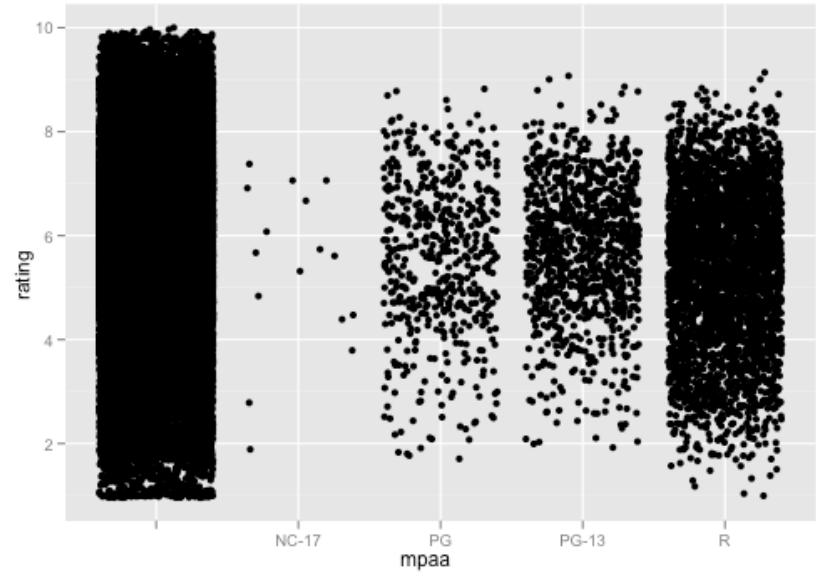
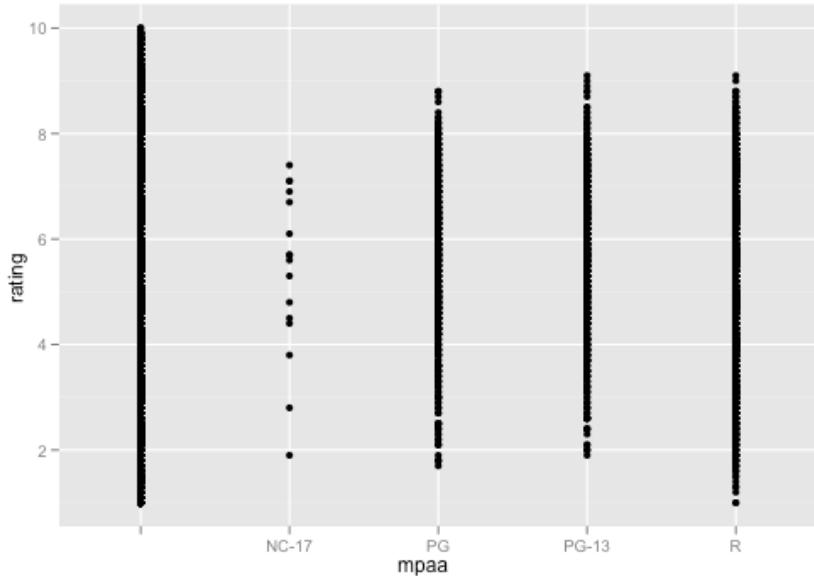
Bump Chart



Library: ggplot2

Overplotting

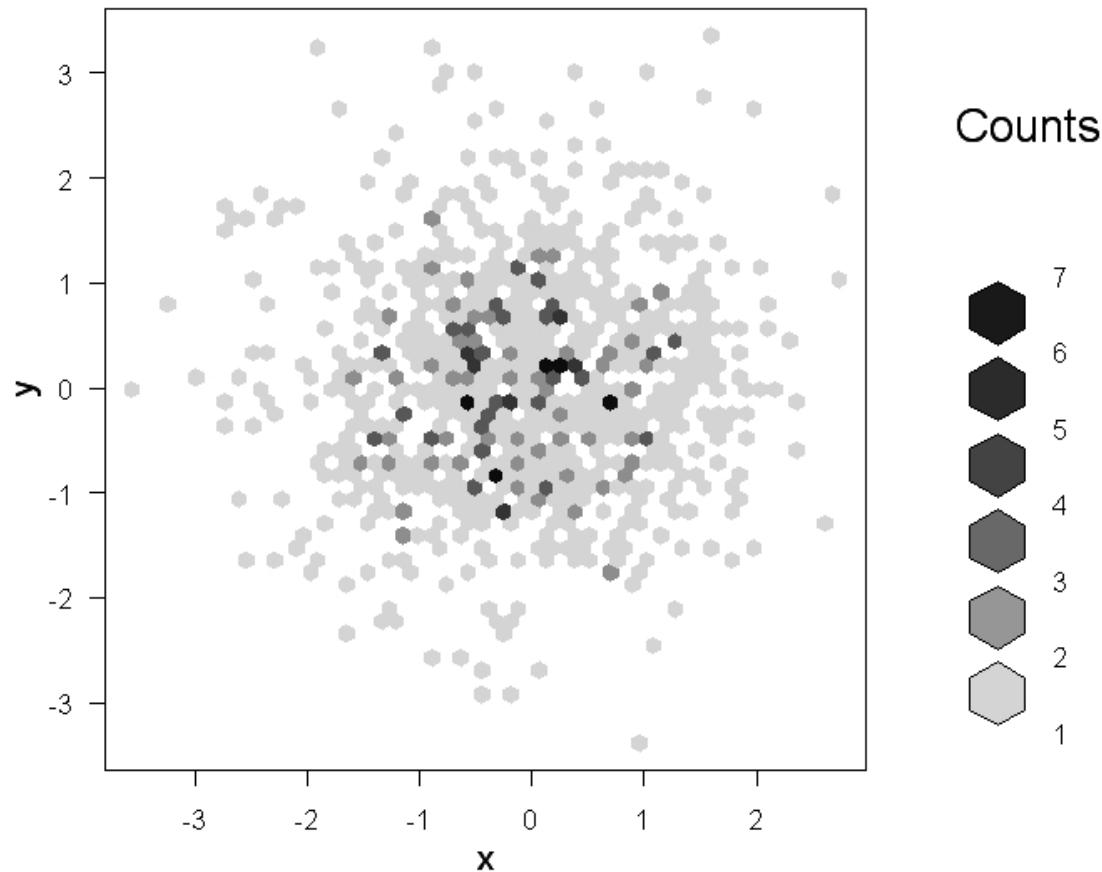
Jittering



Library: ggplot2

Overplotting

Hexagonal Binning



Library: hexbin

