Basic Plots

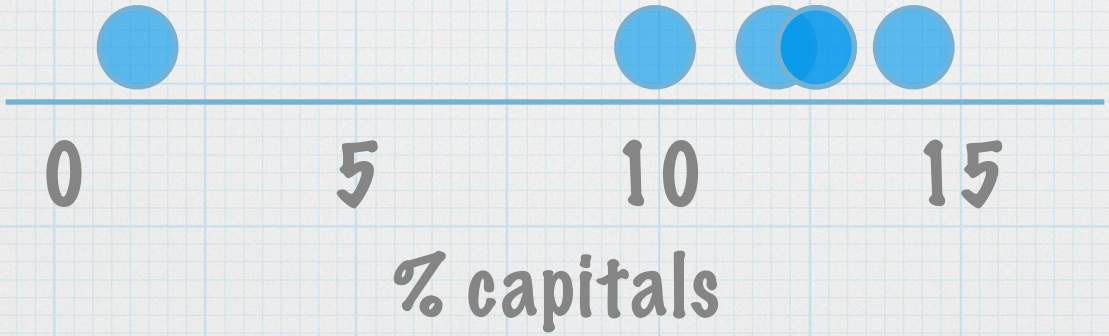
Univariate, bivariate, multivariate
Histogram, boxplot, dotplot, barchart, spine plot
Scatterplot, density plots, mosaic plot
Parallel coordinate plot, profile plots
Maps
Time series plots

Data and its shapes

* Data comes in a lot of different formats
* We will assume that we can always get it into a shape (spread-sheet like) with
  * headers at the top
  * columns for each piece of information
  * and rows for each object

Univariate

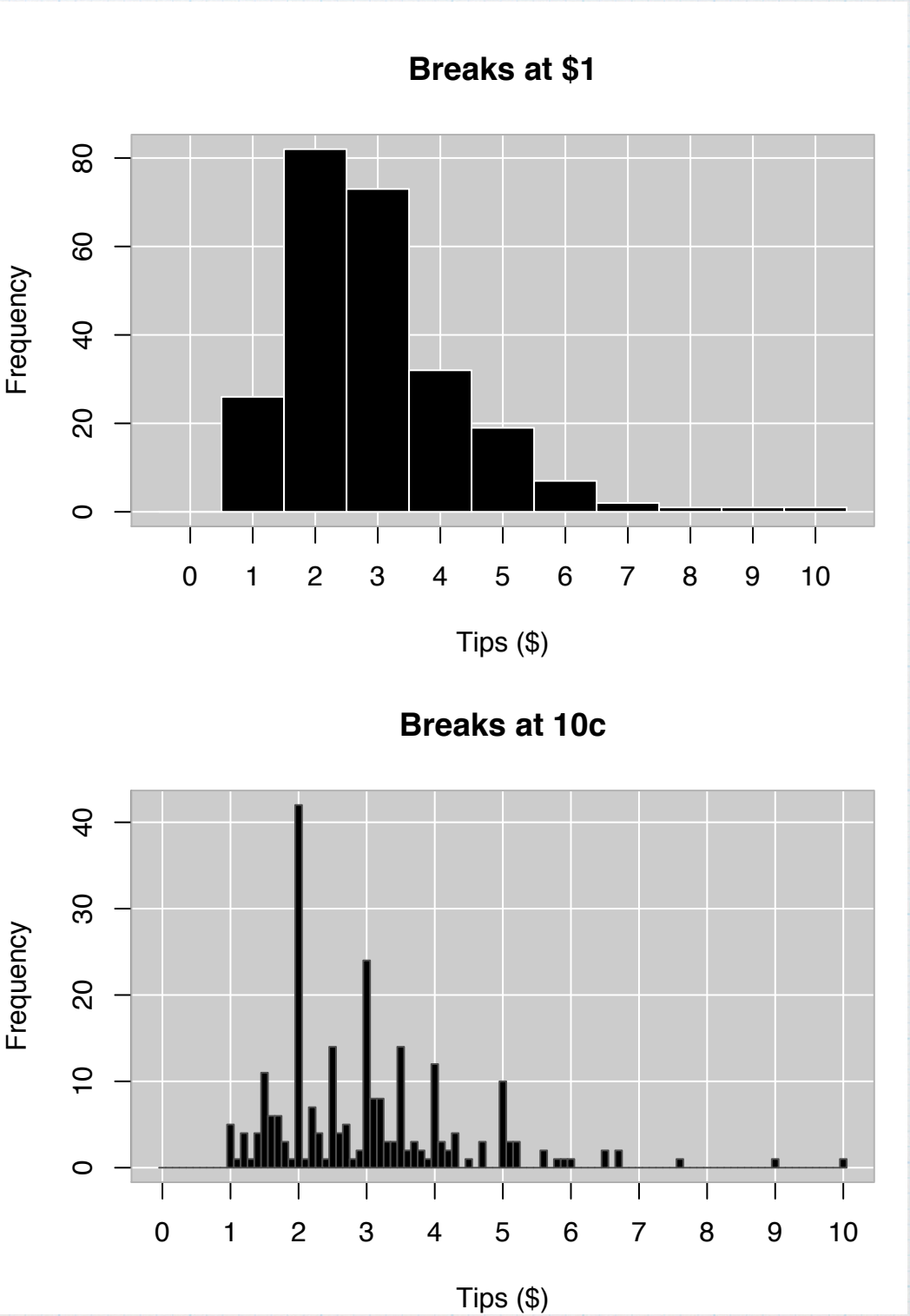
* A dotplot is used for real-valued variables.
* A dot is positioned along an axis to represent the data value.



%Cap
14
12.8
13
1.3
13
10

Univariate

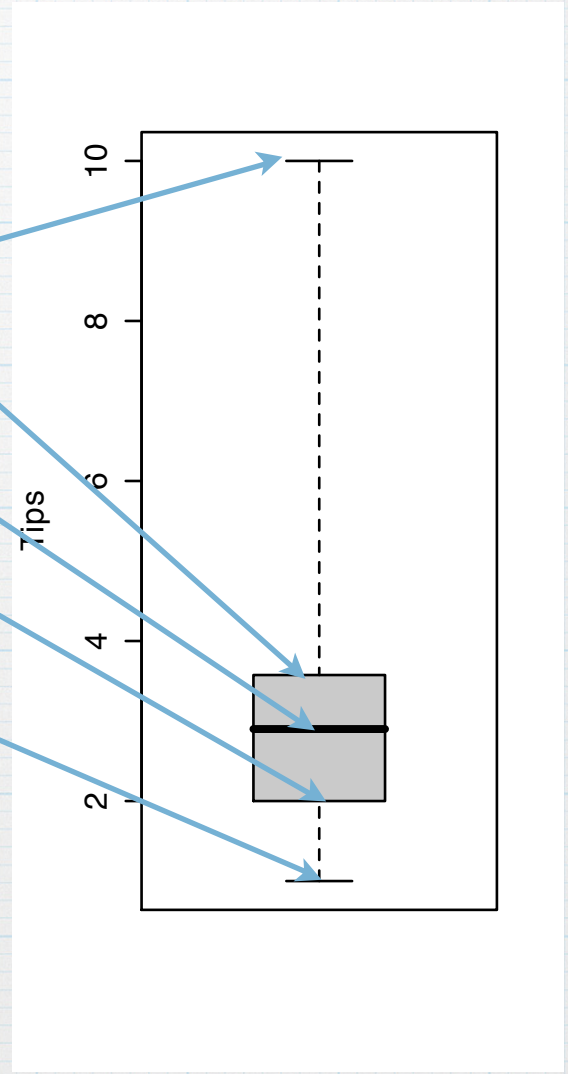
* Histograms are used for real-valued variables.
* Values are binned and the count is displayed by a rectangle



Univariate

- Boxplots are used for real-valued variables
- The data values are summarized by 5 numbers: min, Q1, median, Q3, max
- The boxplot displays just these 5 numbers

Max	10
Q3	3.6
Median	2.9
Q1	2
Min	1



Univariate

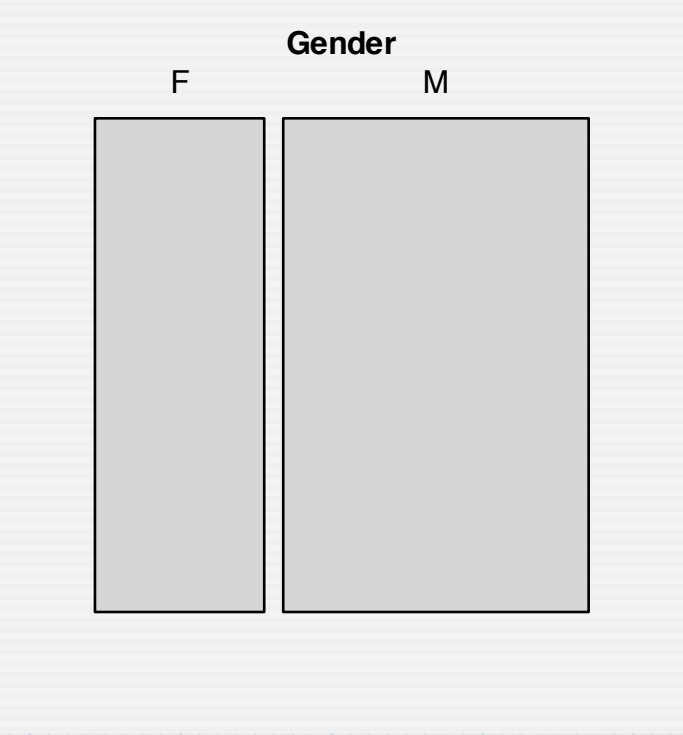
F	M
87	157

* Barcharts are used for categorical data
* The count for each category is represented by the height of a rectangle

Univariate

F	M
87	157

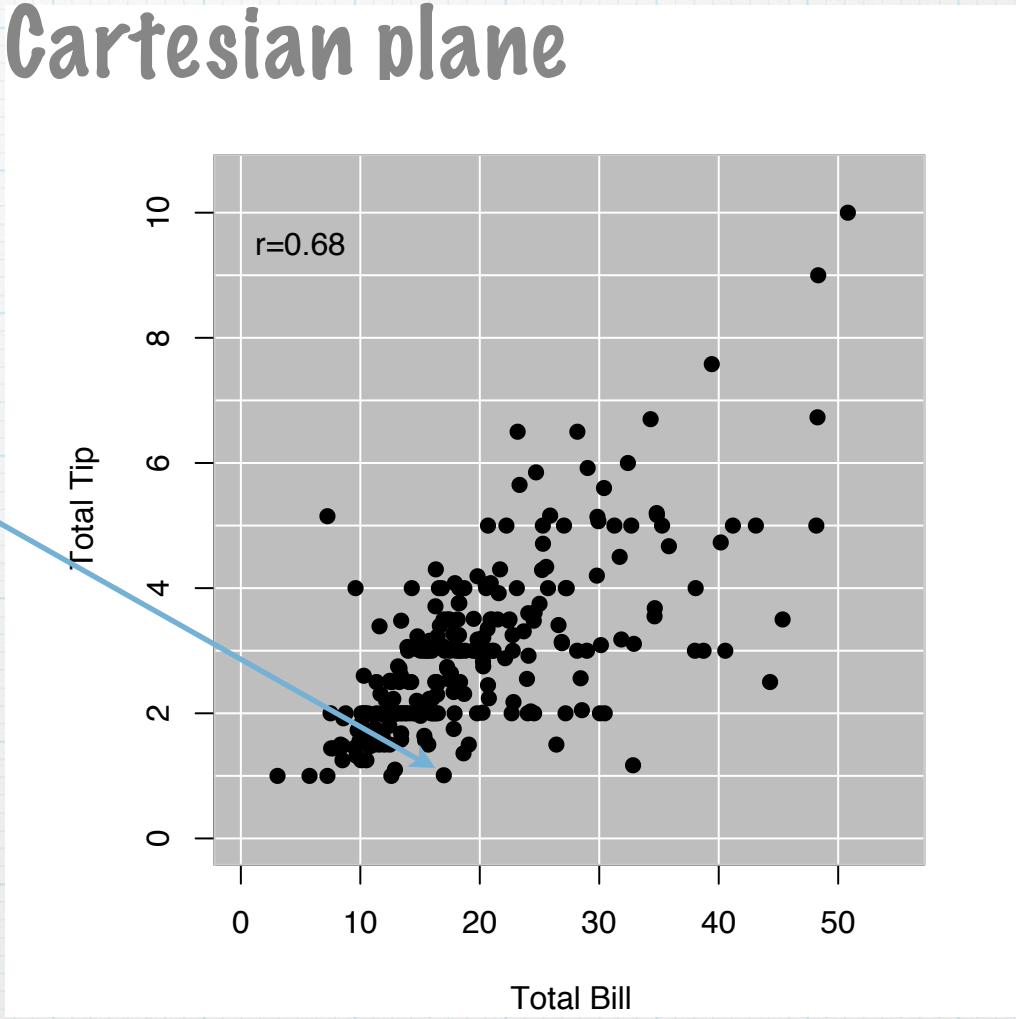
- Spine plots are used for categorical variables
- The count for each category is represented by the width of a rectangle
- Most useful when there are two variables or more.



Bivariate

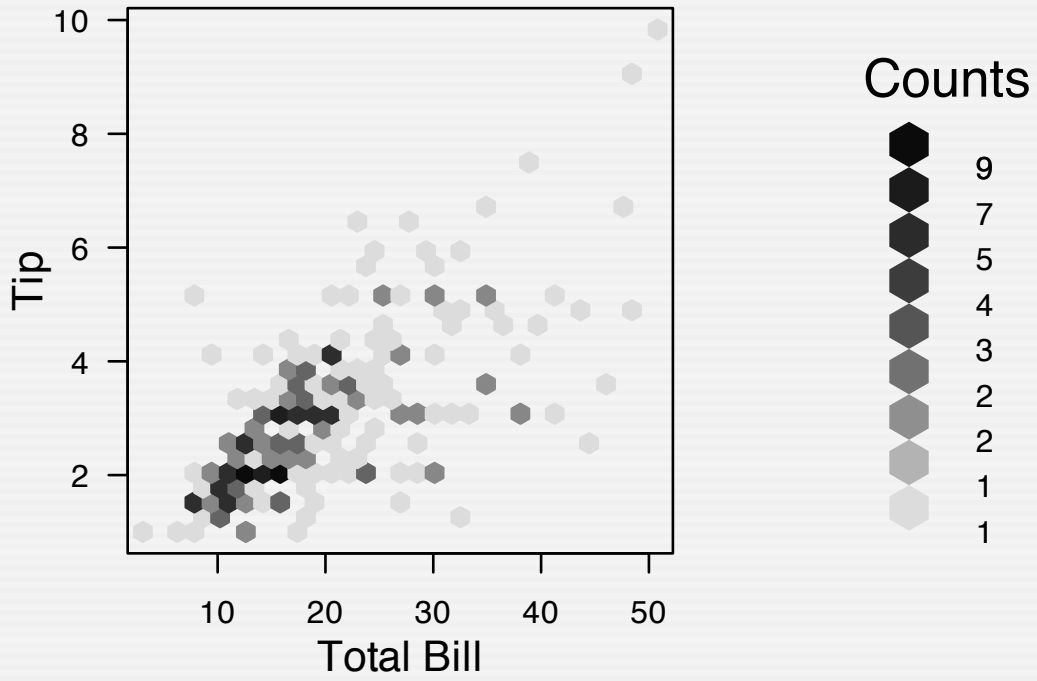
Scatterplots place a dot representing a pair of numbers on a Cartesian plane

Bill	Tip
16.99	1.01
10.34	1.66
21.01	3.5
23.68	3.31
...	...



Bivariate

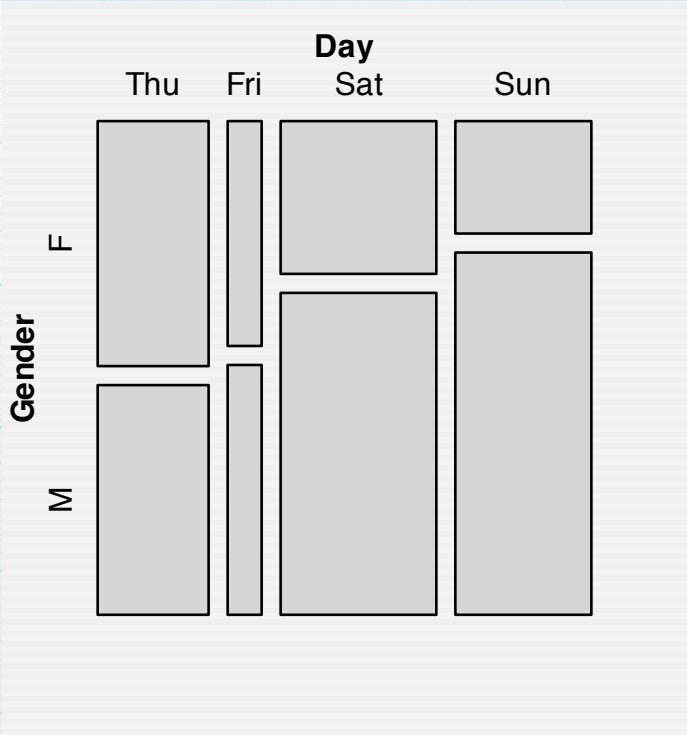
* Density plots : hexagonal grids (Carr)



Bivariate

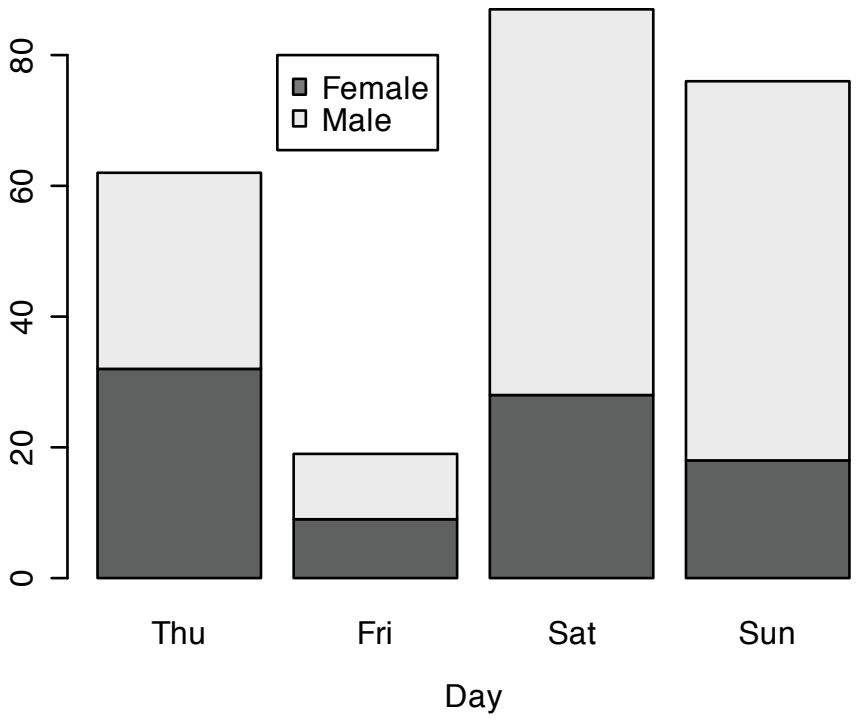
	Thu	Fri	Sat	Sun
F	32	9	28	18
M	30	10	59	58

* A mosaic plot represents a two way table of categorical variables.
* It starts from a spine plot and divides the bars according to counts of a second variable.

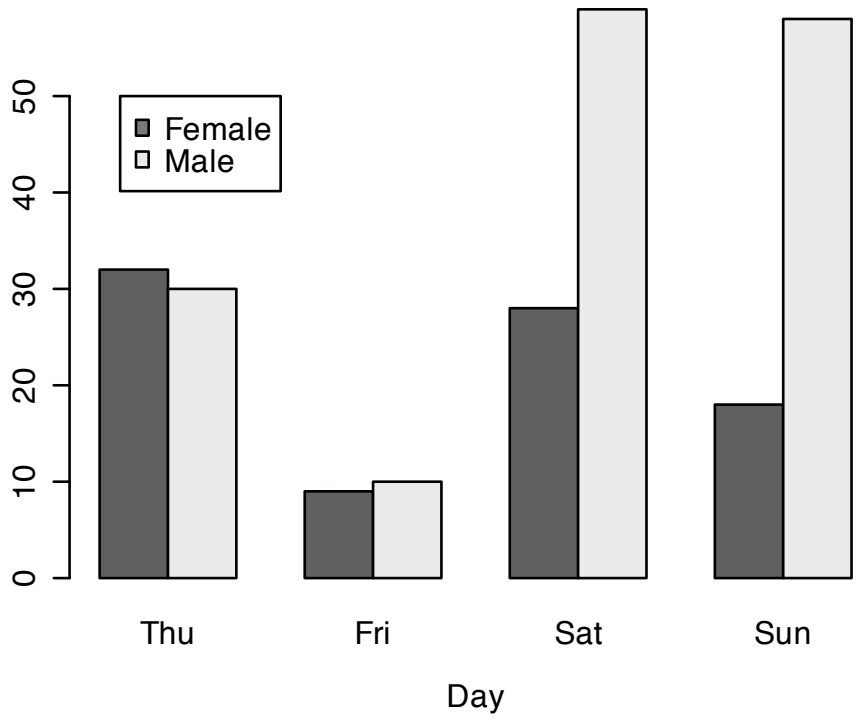


Barcharts with two variables

Stacked barchart



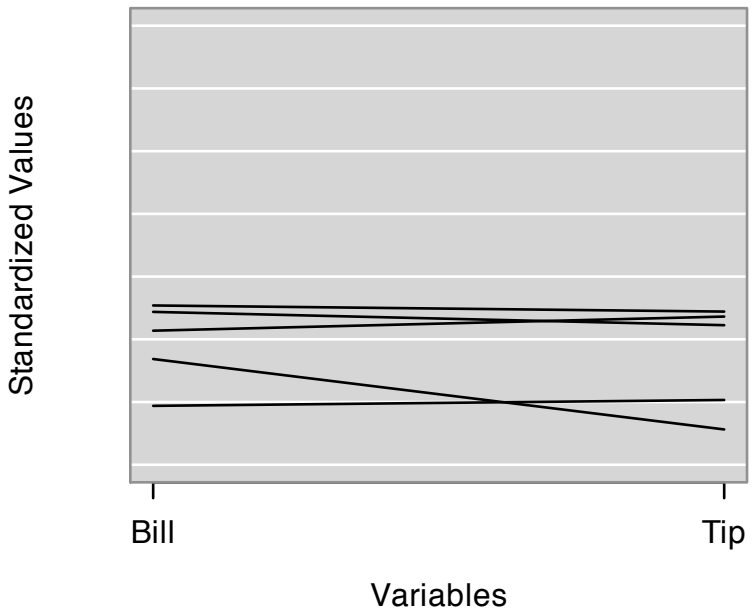
Side-by-side barchart

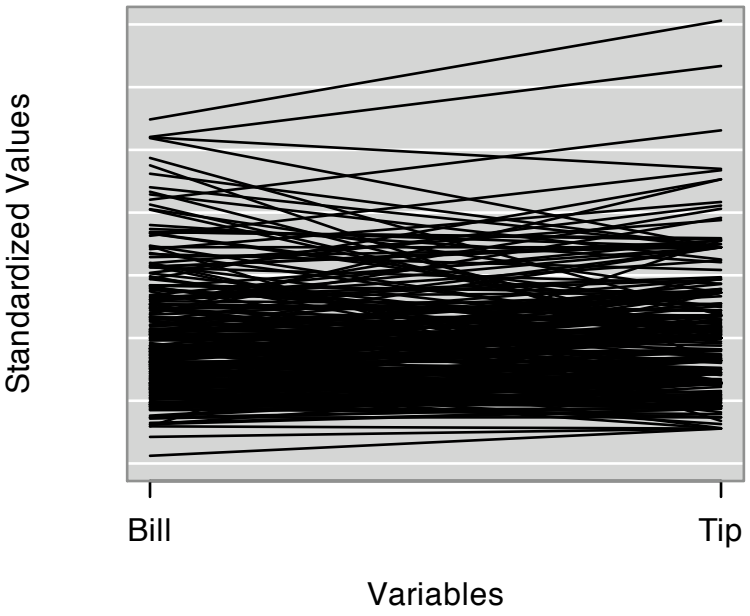


Multivariate

A parallel coordinate plot changes from orthogonal Cartesian axes to parallel axes.

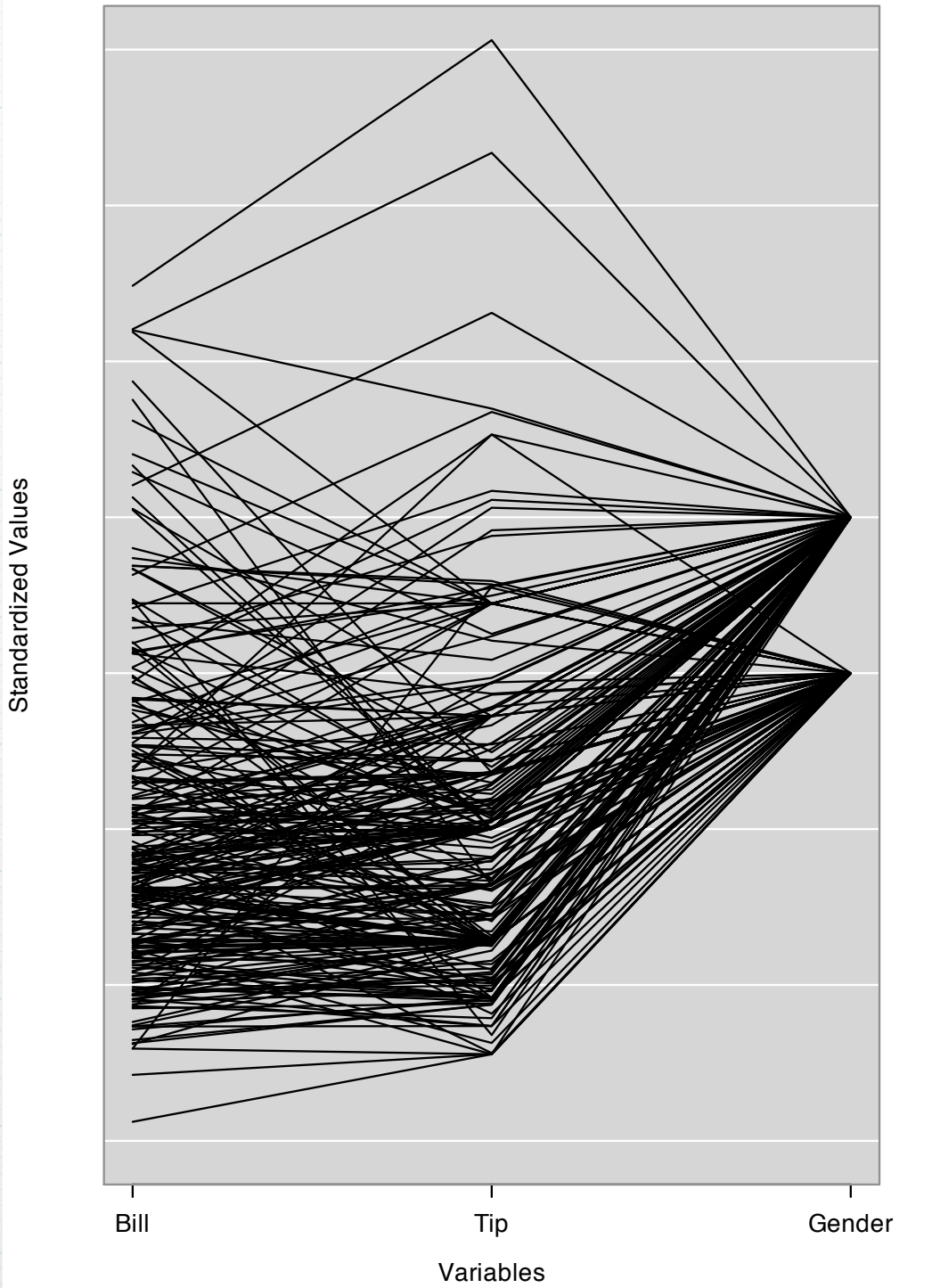
Bill	Tip
16.99	1.01
10.34	1.66
21.01	3.5
23.68	3.31
...	...



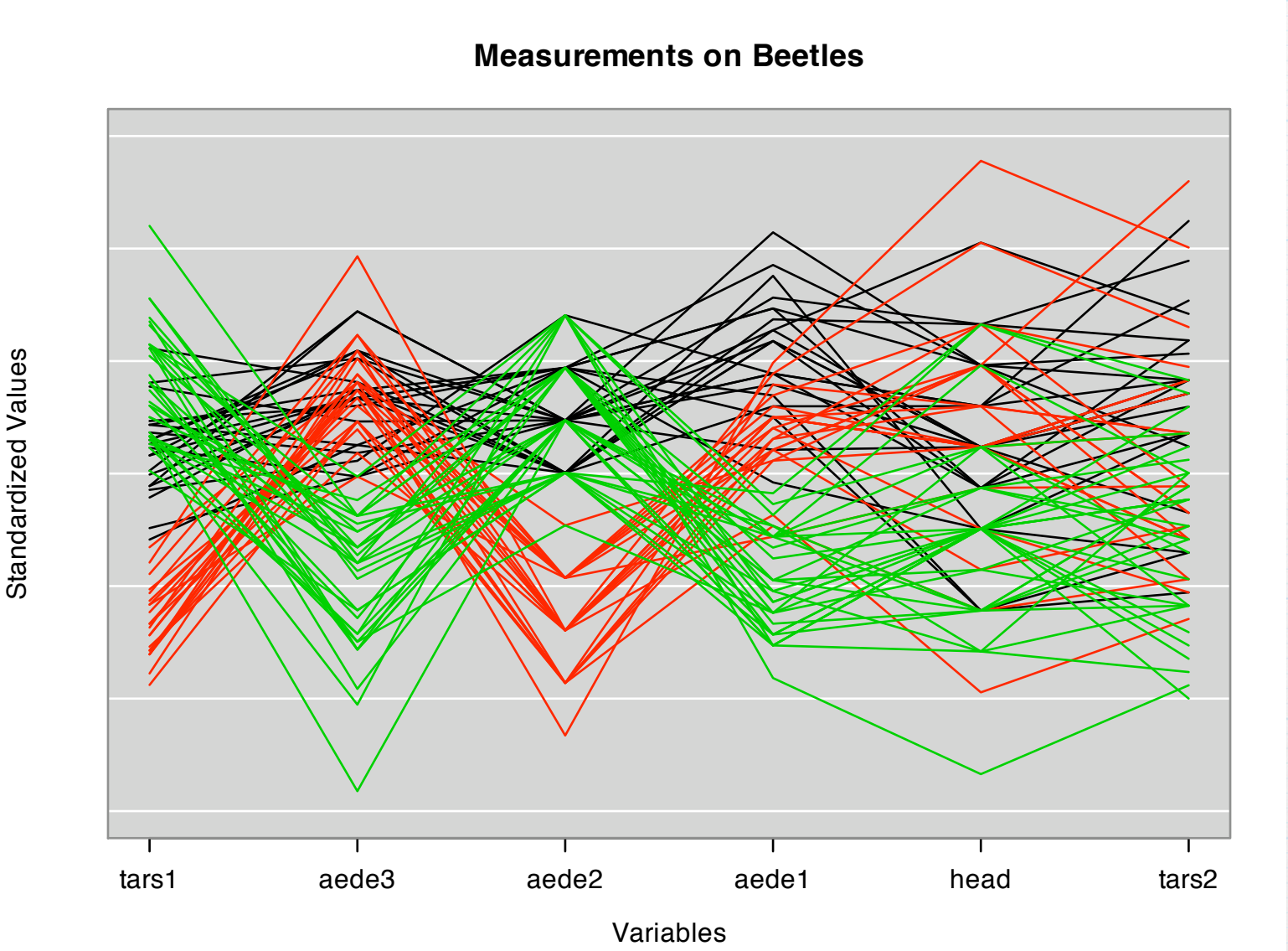


Parallel coordinate plots

Look for patterns in the direction of the lines



Parallel coordinate plots

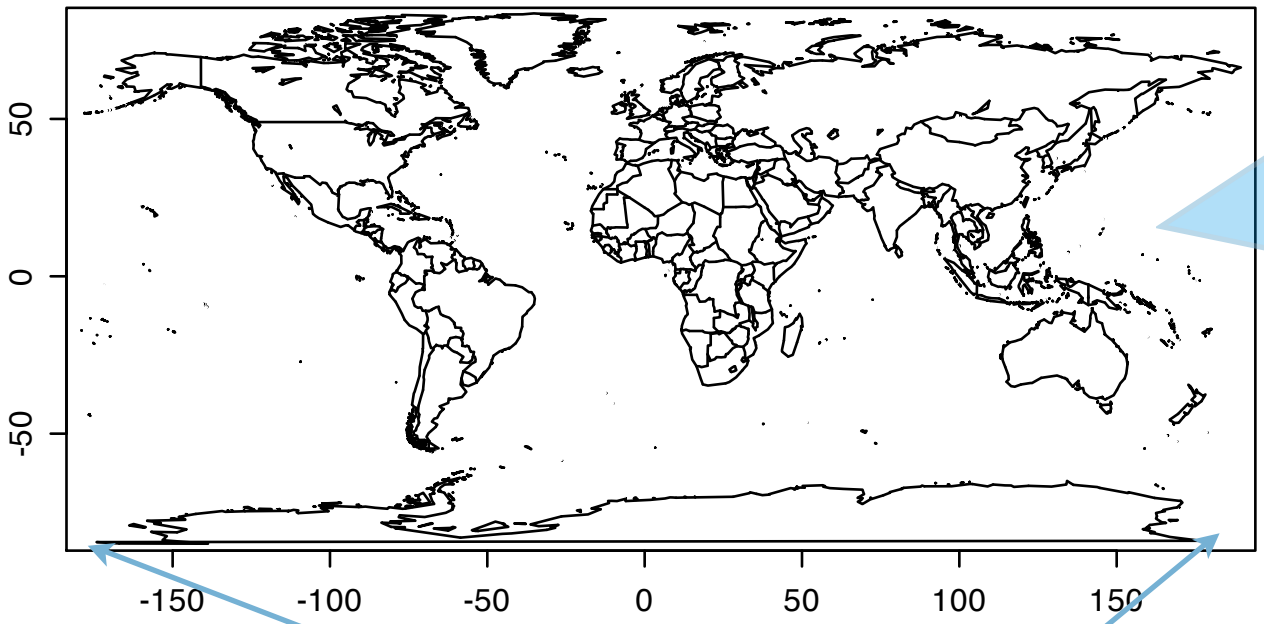


What patterns do you see here?

Maps

* Convention North at top
* The problems with taking longitude at number value
* Aspect ratio of lat to long
* Small regions/areas and reading information

Longitude at numerical value and aspect ratio of longitude to latitude

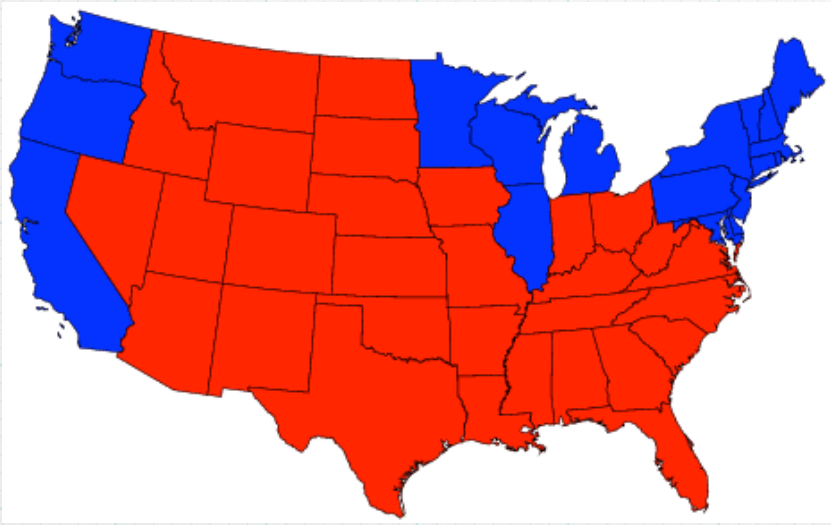


Can you imagine what the world would look like if the vertical and horizontal plot space were equal?

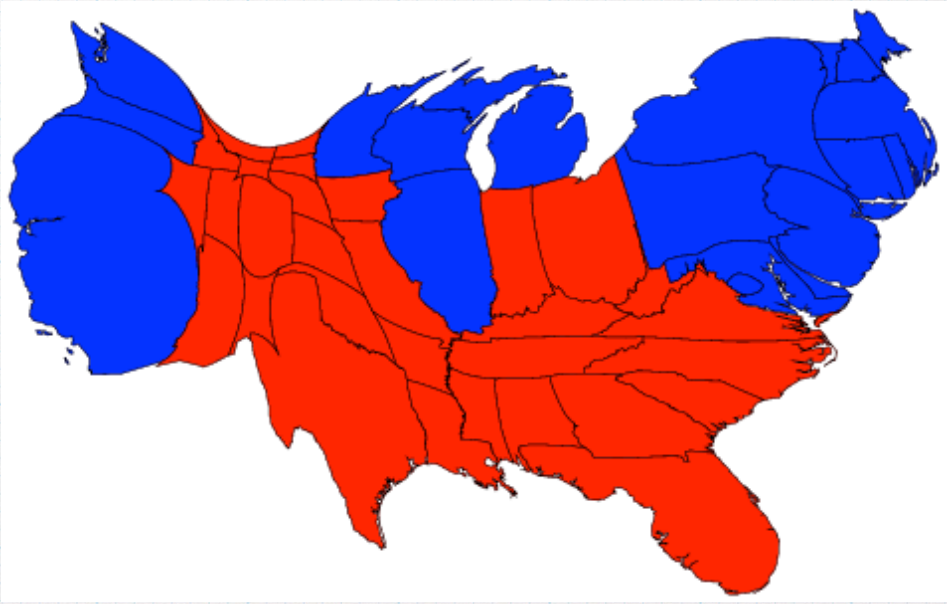
This location is both -180 and 180

Small areas/regions

2004 election results on map: red=republican, blue=democrat



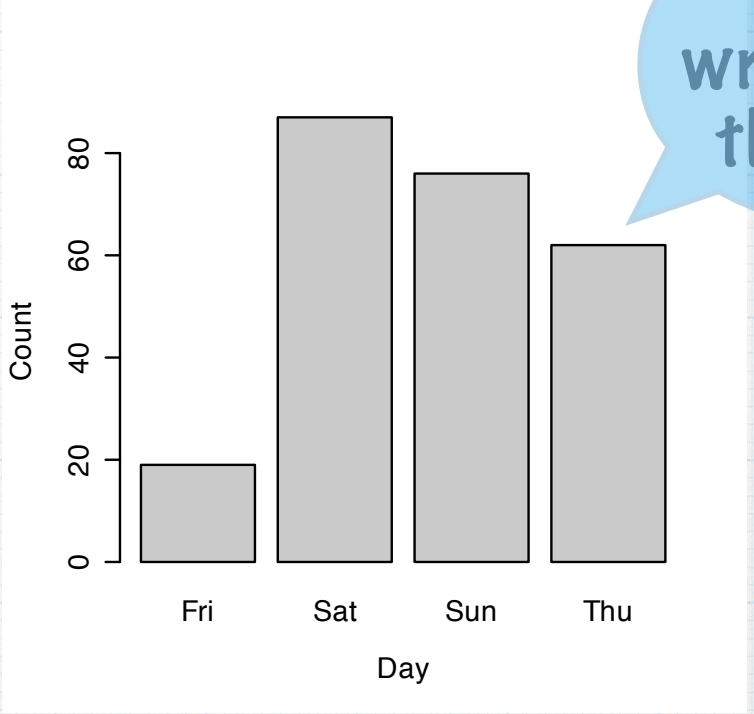
Cartogram of 2004 election results

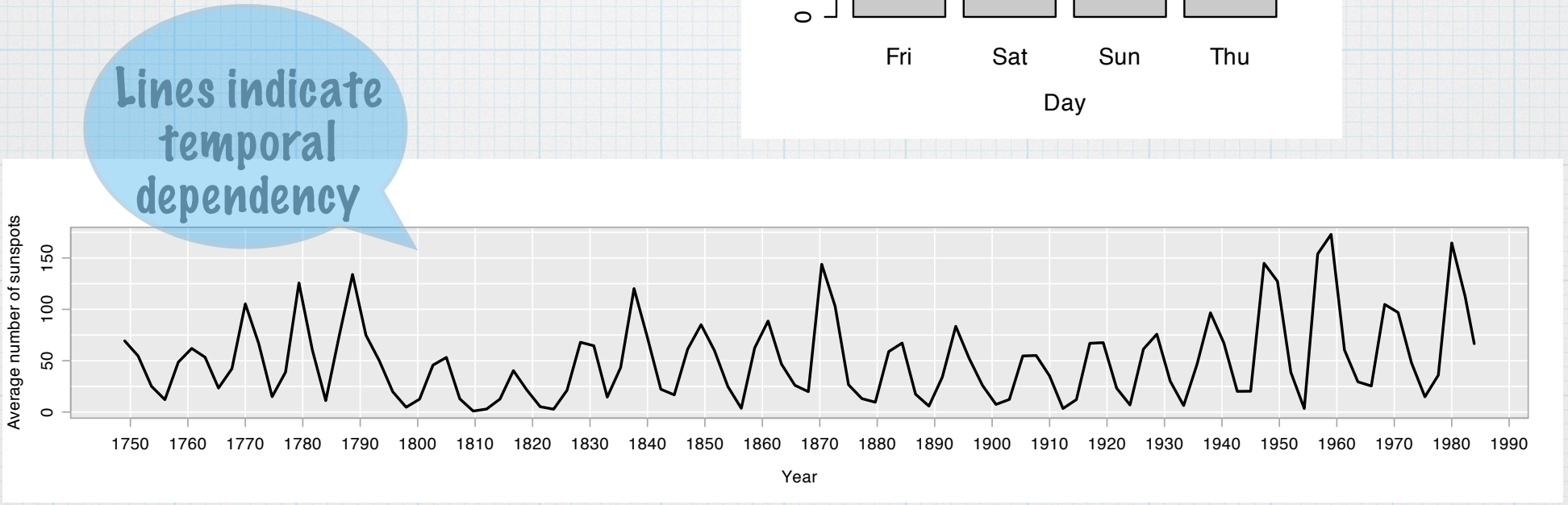


http://www-personal.umich.edu/~mejn/election/

Time series plots

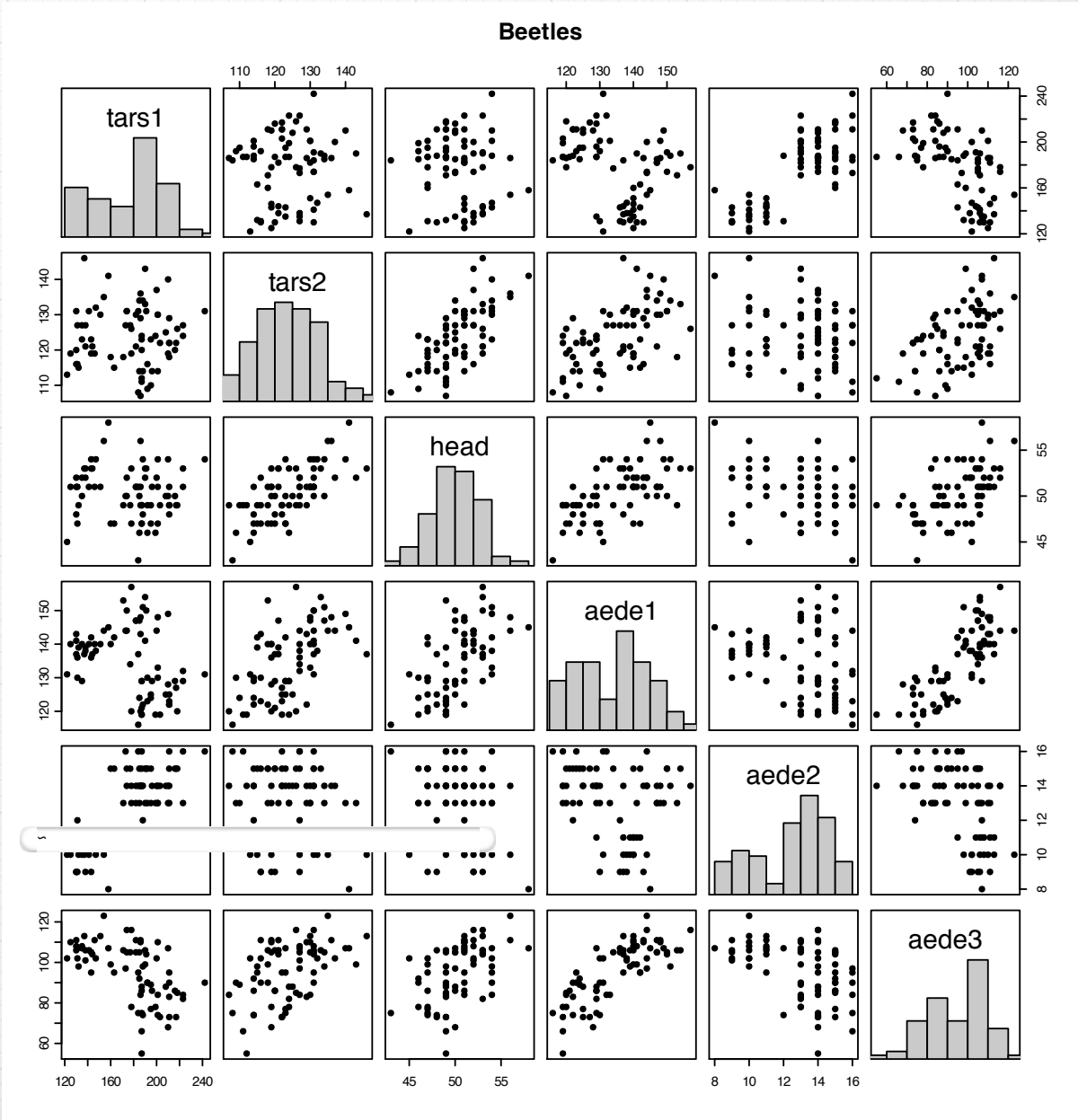
Temporal scale: Days of weeks need to be in conventional order, lines, ...





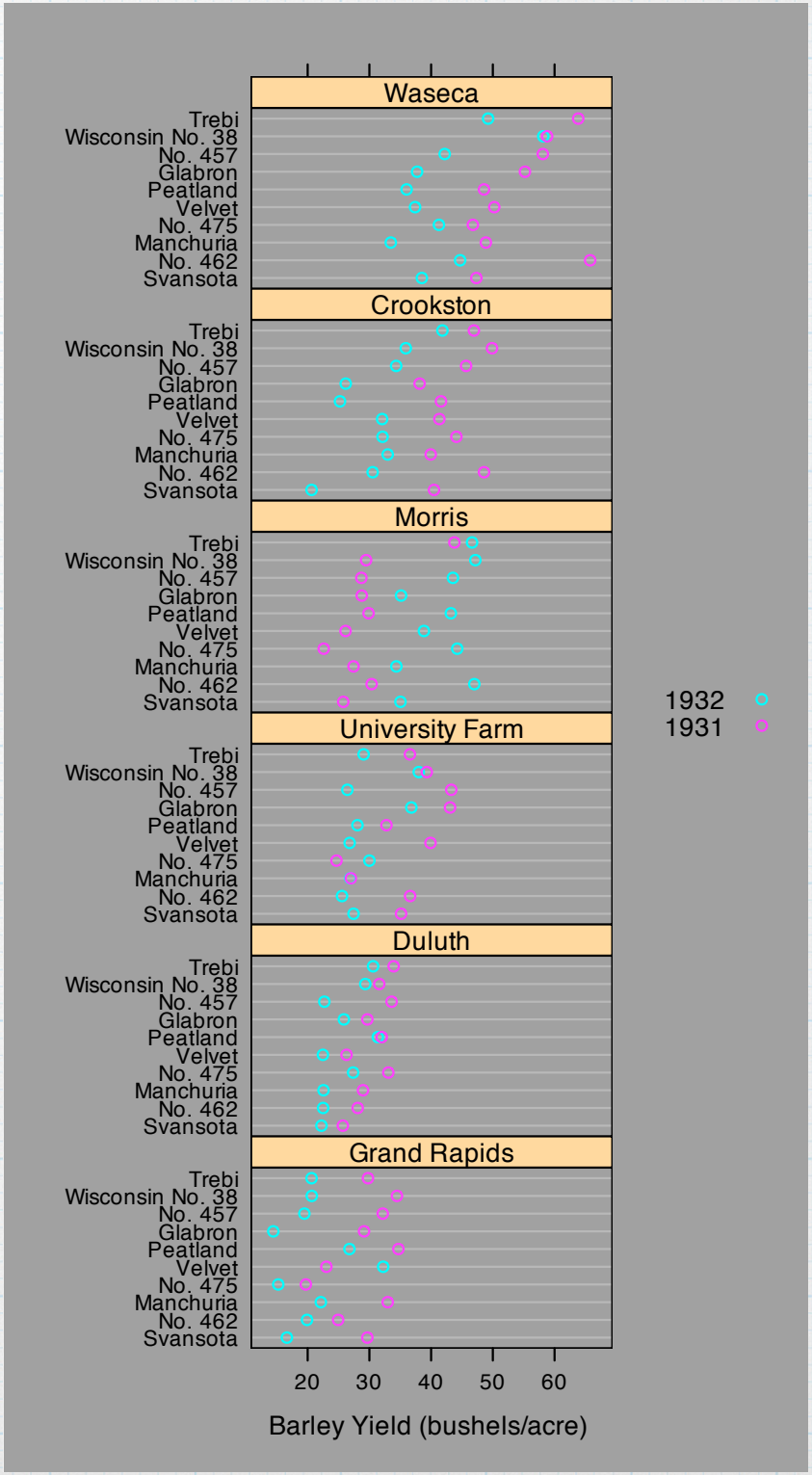
Combinations

Small multiples is an approach advocated by Tufte to plotting multiple variables in a digestible way. This might be considered as combinations of basic plots.



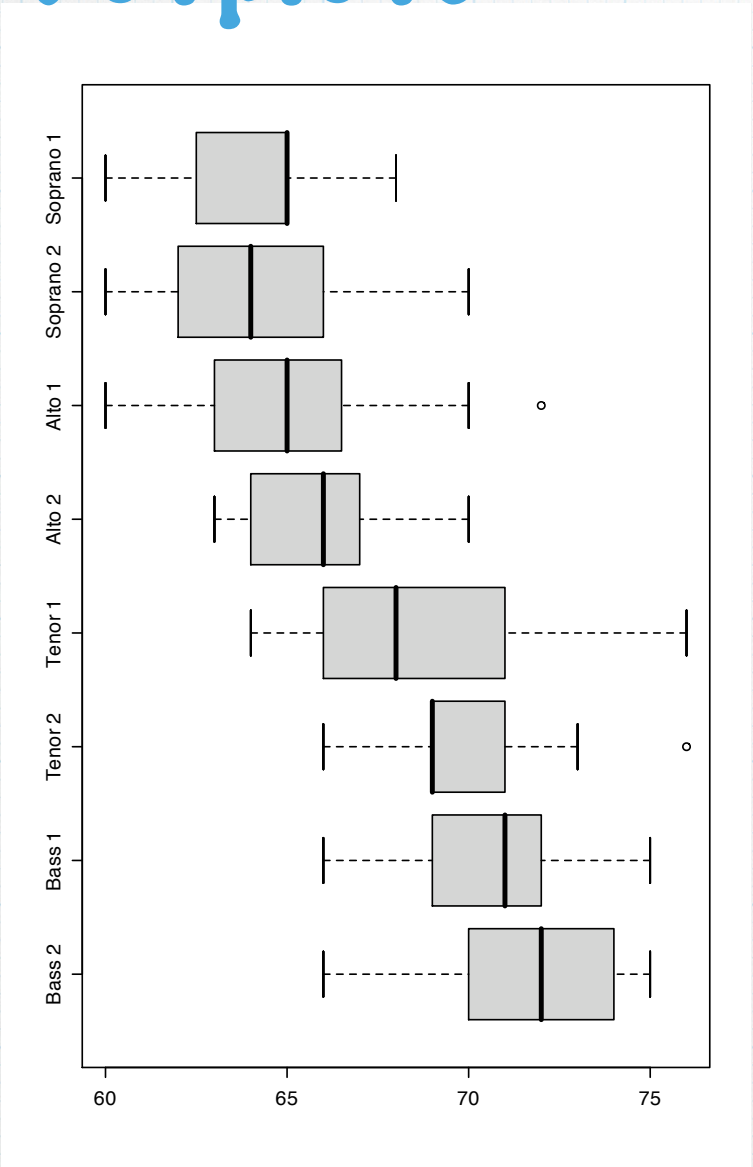
Dotplots

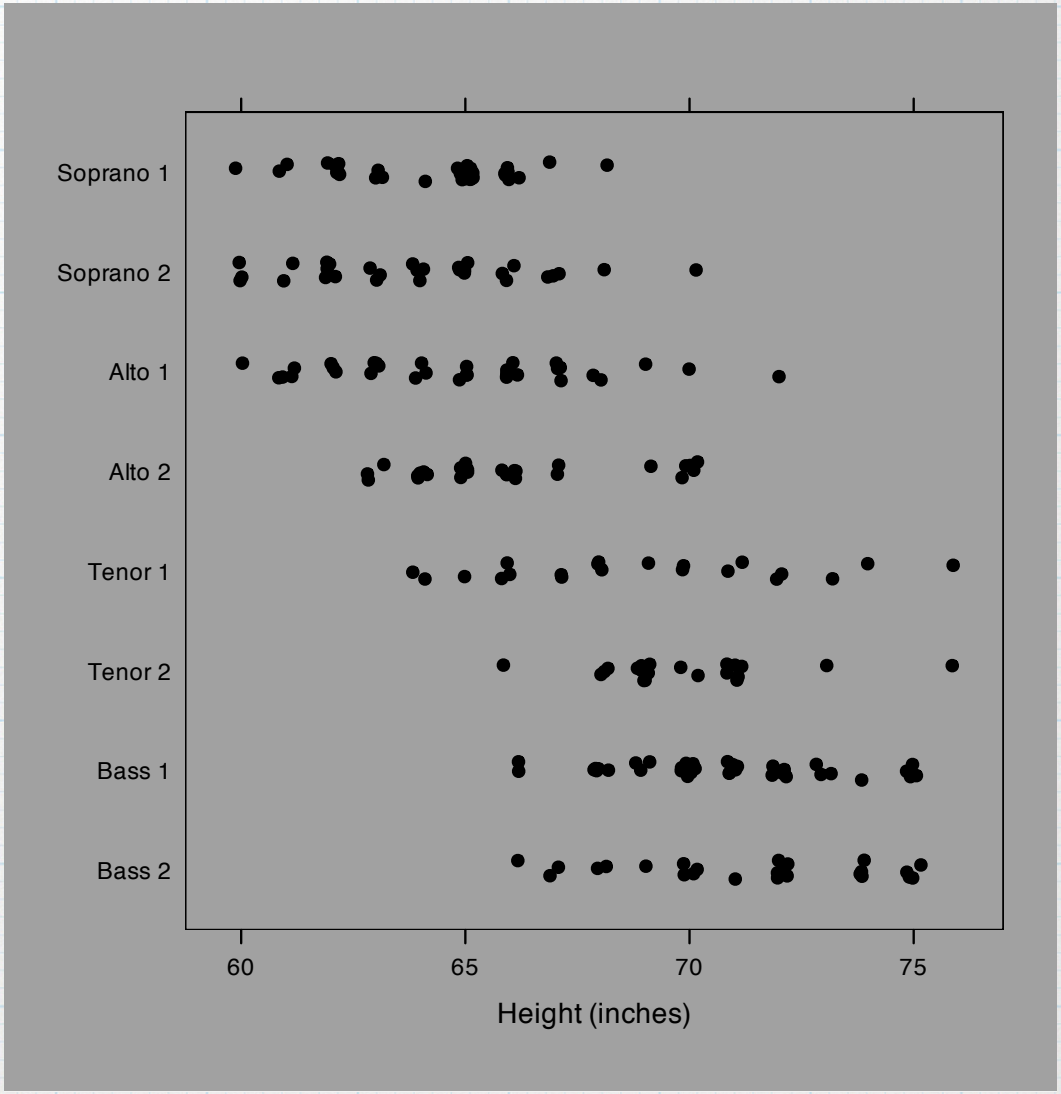
Barley yield for two different years for 6 locations in Minnesota and 10 varieties



Boxplots or Dotplots

Different representations of the heights of the New York Choral Society

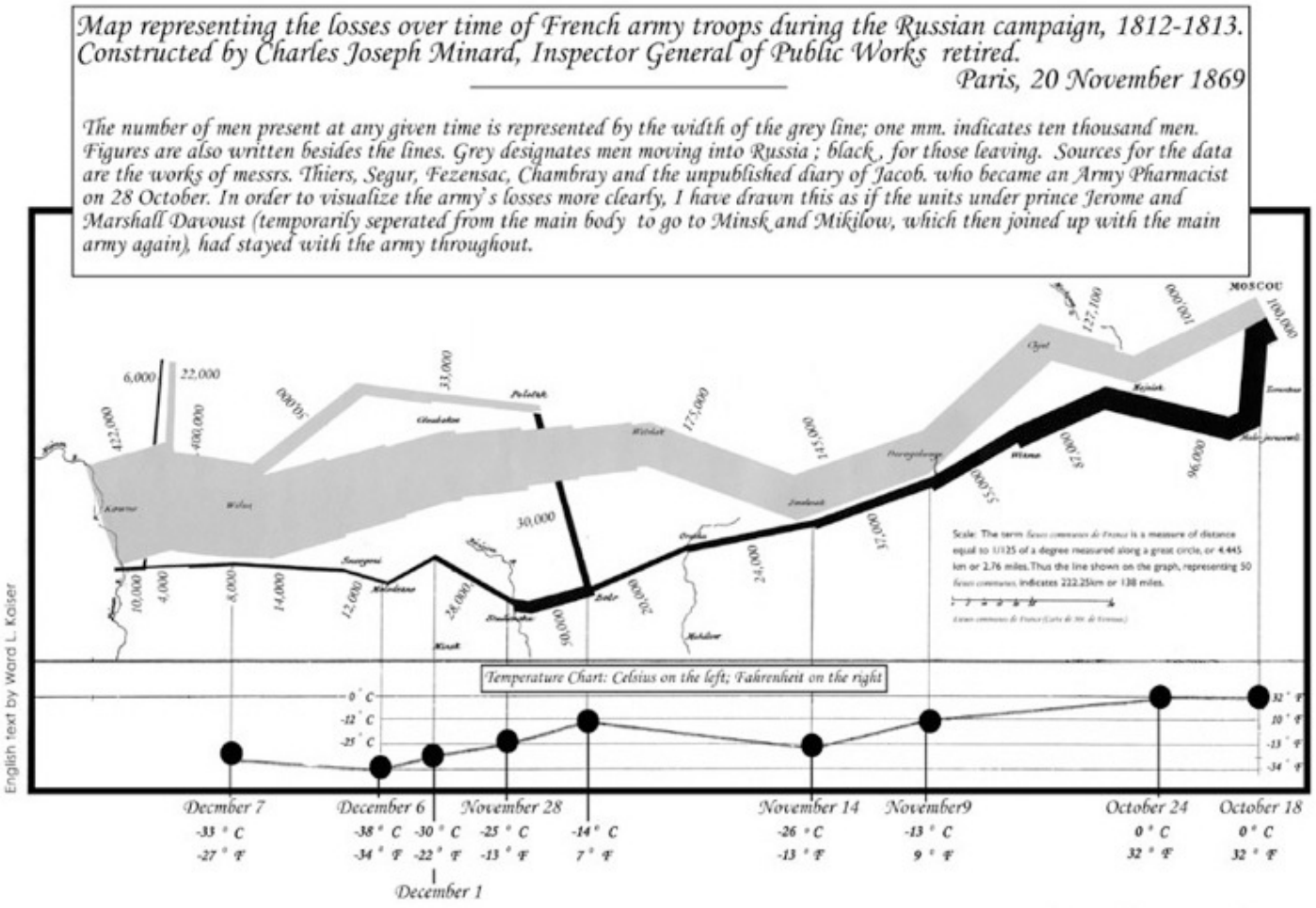




Basic plots form the core

* How does Napoleon's March use basic plots? time series plot + map + barchart
* How is John Snow's cholera map related to a basic plot? Map + scatterplot

Napoleon's March

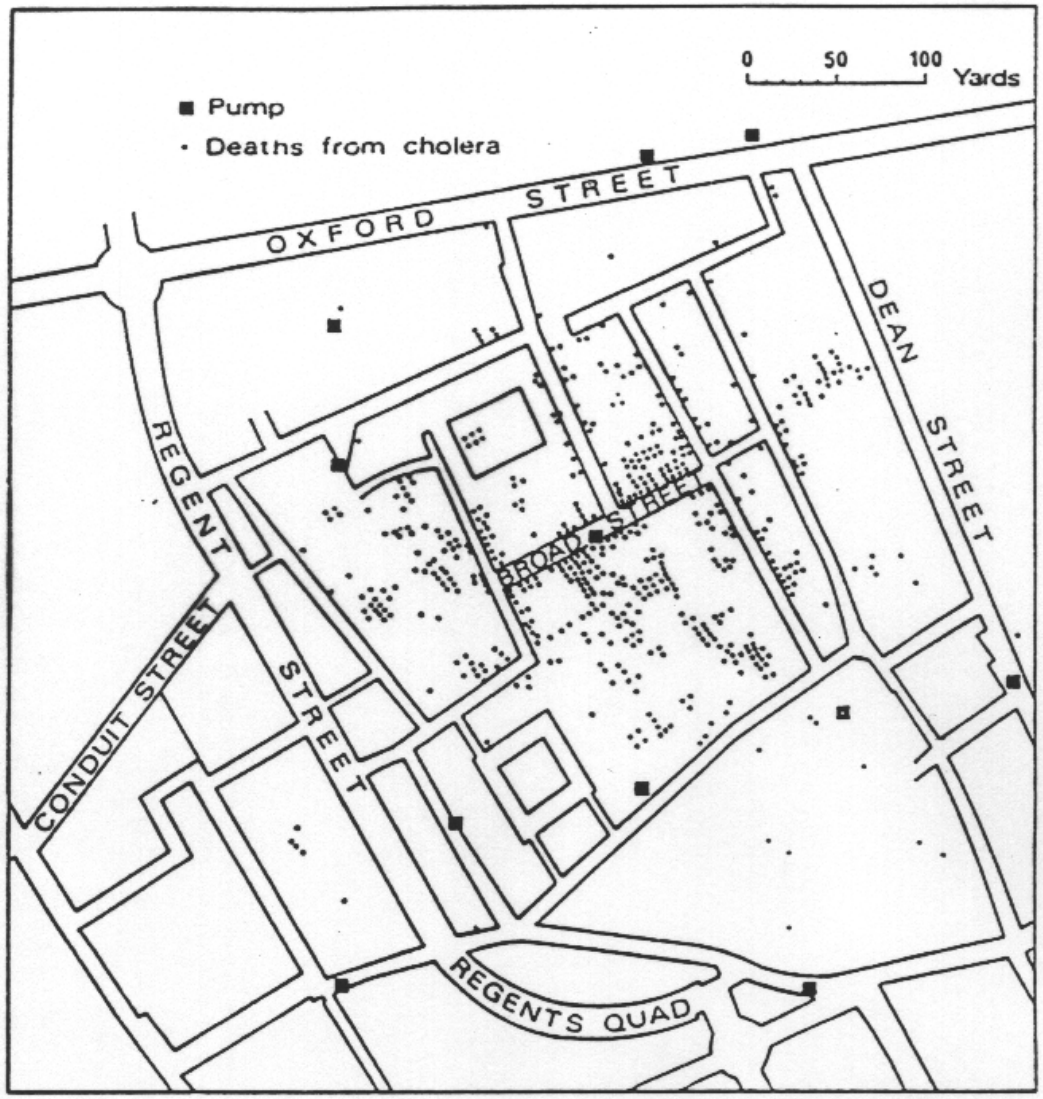


Editor s note: dates & temperatures are only referenced for the retreat from Moscow

Figure 58. ***Minard s map of Napoleon s Russian campaign.***
This graphic has been translated from French to English and modified to most effectively display the temperature data.

Basic plot(s)?

Cholera map

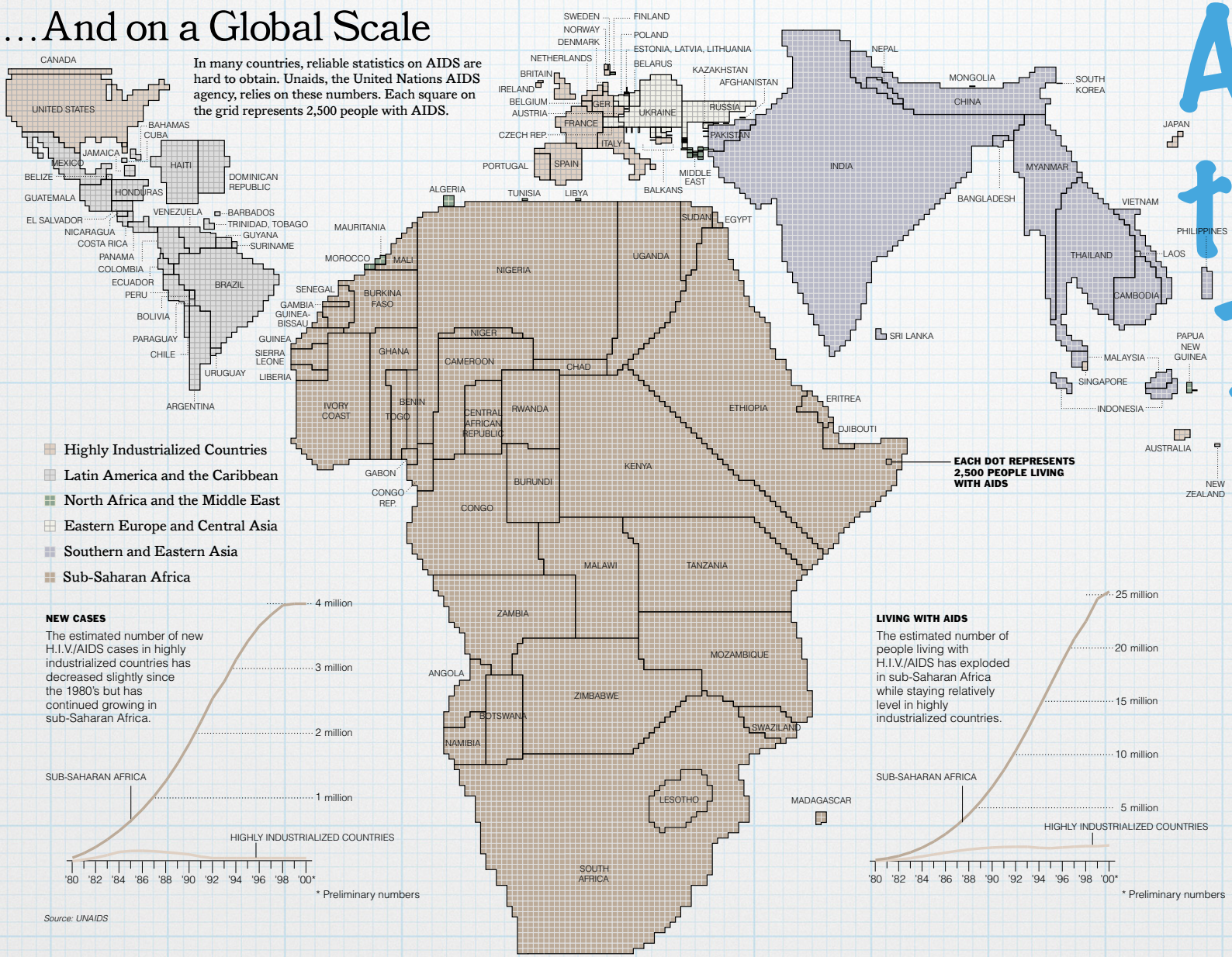


Vignette Figure 1-1. Snow's map of cholera. The affected well is clearly identified by the concentration of cases in its vicinity. Reprinted from Howe, G. M. (1972). *Man, environment, and disease in Britain*. New York: Barnes and Noble Books, p. 178. Copyright 1972. Reproduced by Permission. Original source: Snow, J. (1855). *On the mode of communication of cholera*. London. 1855.

AIDS in the NY Times

…And on a Global Scale

In many countries, reliable statistics on AIDS are hard to obtain. Unaids, the United Nations AIDS agency, relies on these numbers. Each square on the grid represents 2,500 people with AIDS.



- Highly Industrialized Countries
- Latin America and the Caribbean
- North Africa and the Middle East
- Eastern Europe and Central Asia
- Southern and Eastern Asia
- Sub-Saharan Africa

EACH DOT REPRESENTS 2,500 PEOPLE LIVING WITH AIDS

NEW CASES

The estimated number of new H.I.V./AIDS cases in highly industrialized countries has decreased slightly since the 1980's but has continued growing in sub-Saharan Africa.

* Preliminary numbers

LIVING WITH AIDS

The estimated number of people living with H.I.V./AIDS has exploded in sub-Saharan Africa while staying relatively level in highly industrialized countries.

* Preliminary numbers

Source: UNAIDS