Week 5 – Systems and Toolkits Part 2

Yasmin AlNoamany

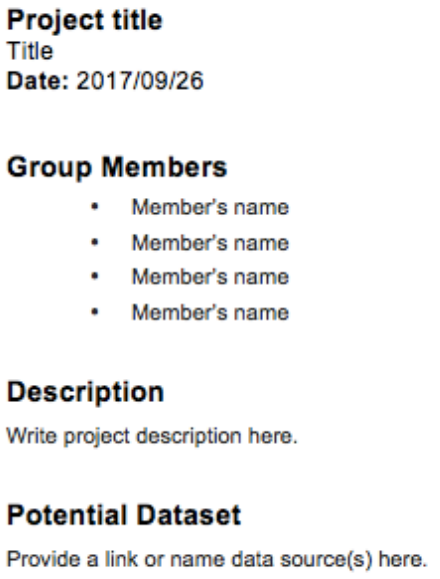
L&S 88-2

University of California, Berkeley

Announcements

- Assignment 3 – Due Oct. 10th

- Semester Project – Link
  - Milestone 1 is due September 29th
  - Elevator pitches on Oct. 3rd
    - Only one of the team will present



Project title
Title
Date: 2017/09/26

Group Members

- Member's name
- Member's name
- Member's name
- Member's name

Description

Write project description here.

Potential Dataset

Provide a link or name data source(s) here.

Communication

- Using bCourses or email me at yasminal@berkeley.edu
  - Put "LS 88-002 WDV" in the subject of your email
  - I may miss your email if you don't do that

- Start working early on assignments
  - I won't reply emails about assignments on the day of submission
  - I may get to you back after 12 – 24 hours in weekdays

- Check out announcements on bCourses
  - If I generate a discussion for Assignment 2 issues, then no emails about the assignment, you should post your question there

- Reminder about late policy:
  - Late assignments lose 1 points for every 24 hours they are late
  - Students should contact the instructor if they have an excuse
  - No assignment will be accepted more than 7 days late

Previous lecture

- Getting Data from the Web
  - Web scrapping
- Choosing Tools to Visualize Data
- Data preprocessing
  - Data Cleaning
- Tools for visualizing data
  - Intro to R and Rstudio

Today's lecture

- Types of data models
- Charts types
- Analyzing and Visualizing data using R and Rstudio
- Introduction to visualization using Tableau

Resources

- Data Analysis and Visualization in R from Data Carpentry lessons
- Producing Simple Graphs with R
- Getting Started with Charts in R
- Get Started with Tableau Desktop tutorial

Objectives

- Explain different types of data models
- Describe different variable types (categories)
- Define metadata
- Know when to use a table versus a graph
- Identify four fundamental types of symbolic displays
- Distinguish different graph types
- Analyze and visualize data using R
- Use Tableau for visualizing data

Data

- Data comes in many different forms
  - Typically, not in the way you want them
- We often characterize data through three components, or objects:
  - Items of interest
    - (students, courses, terms, ...)
  - Attributes
    - Characteristics or properties of data
    - (name, age, GPA, number, date, ...)
  - Relations
    - How two or more objects relate
    - (student takes course, course during term, ...)

Example

- Cars
  - model
  - year
  - miles per gallon
  - cost
  - number of cylinders
  - weights

    …

Data types – Stevens' typology

- Three Main Types of Variables
  - Nominal (qualitative)
    - equal or not equal to other values
    - Example: gender
  - Ordinal (qualitative)
    - obeys < relation, ordered set
    - Example: fr, so, jr, sr
  - Quantitative
    - can do math on them
    - Example: age

Alternate characterization

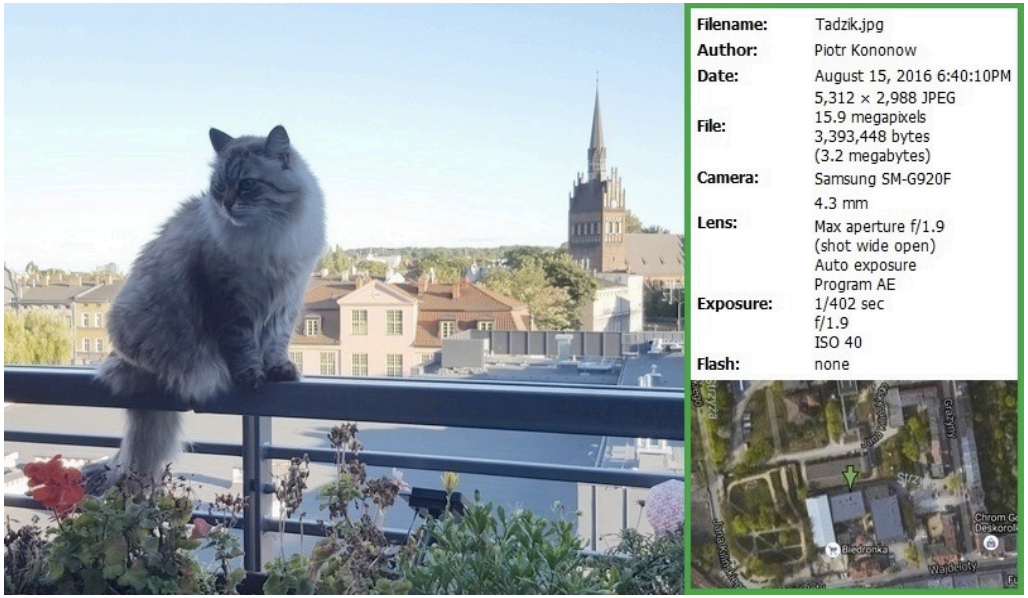
- Two types of data
  - Quantitative/numerical − numerical information
  - Examples:
    - age, height
  - Categorical/qualitative - How attributes relate to each other Intervals
    - Nominal − male, female
    - Ordinal − fr, so, jr, sr
    - Interval − 0-50, 50-100, etc.

Can you identify the type of data that appears below?

- Students home cities
- Weight
- Months
- Year

Metadata

- Data about data
  - Descriptive information about the data
- Might be something as simple as the type of a variable, or could be more complex
- For times when the table itself just isn't enough



https://dataedo.com/blog/what-is-metadata-examples

Data representation

S. Few
Show Me the Numbers

- Tables
  - Data is expressed in the form of text (that is, words and numbers, rather than graphically)
  - Data is arranged in columns and rows
  - Tables work best when the display will be used to look up individual values or the quantitative values must be precise.

- Graphs
  - Data is expressed graphically (that is, as a picture)
  - Data is displayed in relation to one or more axes along which run scales that assign meaning to the values
  - Graphs work best when the message you wish to communicate resides in the shape of the data

Example

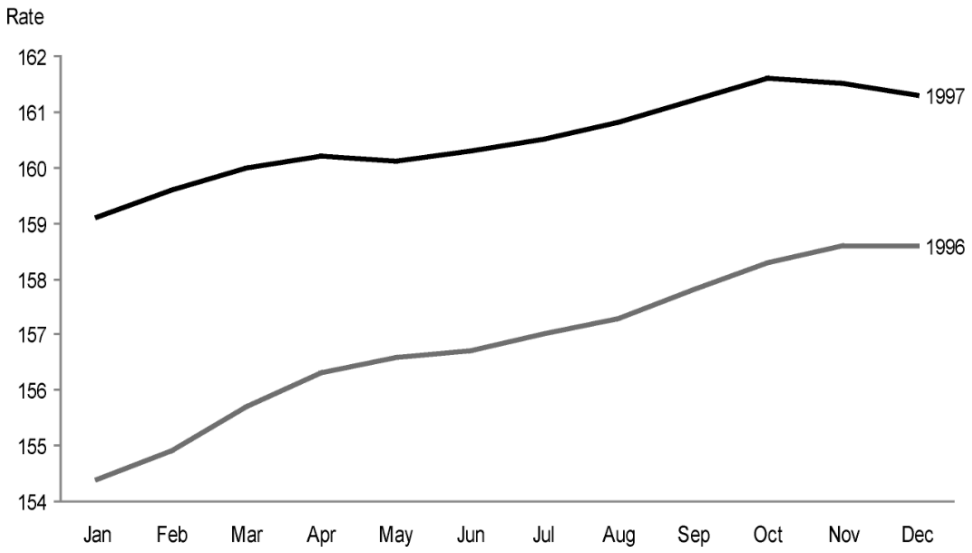
S. Few
Show Me the Numbers

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
1990	127.4	128.0	128.7	128.9	129.2	129.9	130.4	131.6	132.7	133.5	133.8	133.8	130.7
1991	134.6	134.8	135.0	135.2	135.6	136.0	136.2	136.6	137.2	137.4	137.8	137.9	136.2
1992	138.1	138.6	139.3	139.5	139.7	140.2	140.5	140.9	141.3	141.8	142.0	141.9	140.3
1993	142.6	143.1	143.6	144.0	144.2	144.4	144.4	144.8	145.1	145.7	145.8	145.8	144.5
1994	146.2	146.7	147.2	147.4	147.5	148.0	148.4	149.0	149.4	149.5	149.7	149.7	148.2
1995	150.3	150.9	151.4	151.9	152.2	152.5	152.5	152.9	153.2	153.7	153.6	153.5	152.4
1996	154.4	154.9	155.7	156.3	156.6	156.7	157.0	157.3	157.8	158.3	158.6	158.6	156.9
1997	159.1	159.6	160.0	160.2	160.1	160.3	160.5	160.8	161.2	161.6	161.5	161.3	160.5
1998	161.6	161.9	162.2	162.5	162.8	163.0	163.2	163.4	163.6	164.0	164.0	163.9	163.0
1999	164.3	164.5	165.0	166.2	166.2	166.2	166.7	167.1	167.9	168.2	168.3	168.3	166.6
2000	168.8	169.8	171.2	171.3	171.5	172.4	172.8	172.8	173.7	174.0	174.1	174.0	172.2
2001	175.1	175.8	176.2	176.9	177.7	178.0	177.5	177.5	178.3	177.7	177.4	176.7	177.1
2002	177.1	177.8	178.8	179.8	179.8	179.9	180.1	180.7	181.0	181.3	181.3	180.9	179.9

Example

S. Few
Show Me the Numbers

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
1990	127.4	128.0	128.7	128.9	129.2	129.9	130.4	131.6	132.7	133.5	133.8	133.8	130.7
1991	134.6	134.8	135.0	135.2	135.6	136.0	136.2	136.6	137.2	137.4	137.8	137.9	136.2
1992	138.1	138.6	139.3	139.5	139.7	140.2	140.5	140.9	141.3	141.8	142.0	141.9	140.3
1993	142.6	143.1	143.6	144.0	144.2	144.4	144.4	144.8	145.1	145.7	145.8	145.8	144.5
1994	146.2	146.7	147.2	147.4	147.5	148.0	148.4	149.0	149.4	149.5	149.7	149.7	148.2
1995	150.3	150.9	151.4	151.9	152.2	152.5	152.5	152.9	153.2	153.7	153.6	153.5	152.4
1996	154.4	154.9	155.7	156.3	156.6	156.7	157.0	157.3	157.8	158.3	158.6	158.6	156.9
1997	159.1	159.6	160.0	160.2	160.1	160.3	160.5	160.8	161.2	161.6	161.5	161.3	160.5
1998	161.6	161.9	162.2	162.5	162.8	163.0	163.2	163.4	163.6	164.0	164.0	163.9	163.0
1999	164.3	164.5	165.0	166.2	166.2	166.2	166.7	167.1	167.9	168.2	168.3	168.3	166.6
2000	168.8	169.8	171.2	171.3	171.5	172.4	172.8	172.8	173.7	174.0	174.1	174.0	172.2
2001	175.1	175.8	176.2	176.9	177.7	178.0	177.5	177.5	178.3	177.7	177.4	176.7	177.1
2002	177.1	177.8	178.8	179.8	179.8	179.9	180.1	180.7	181.0	181.3	181.3	180.9	179.9



Data tables

- We take raw data and transform it into a model/form that is more workable
- Main idea:
  - Individual items are called cases
  - Cases have variables (attributes)

Data table format

	$Case_1$	$Case_2$	$Case_3$	...
$Variable_1$	$Value_{11}$	$Value_{21}$	$Value_{31}$	
$Variable_2$	$Value_{12}$	$Value_{22}$	$Value_{32}$	
$Variable_3$	$Value_{13}$	$Value_{23}$	$Value_{33}$	
...				

Example

	Mary	Jim	Sally	Mitch	...
SSN	145	294	563	823	
Age	23	17	47	29	
Hair	brown	black	blonde	red	
GPA	2.9	3.7	3.4	2.1	
...					

People in class

Basic symbolic displays

- Graphs
- Charts
- Maps
- Diagrams

S. Kosslyn, “Understanding charts and graphs”, Applied Cognitive Psychology, 1989.

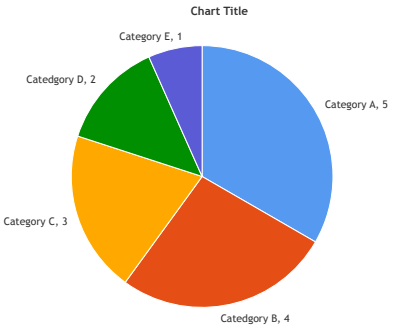
Basic symbolic displays

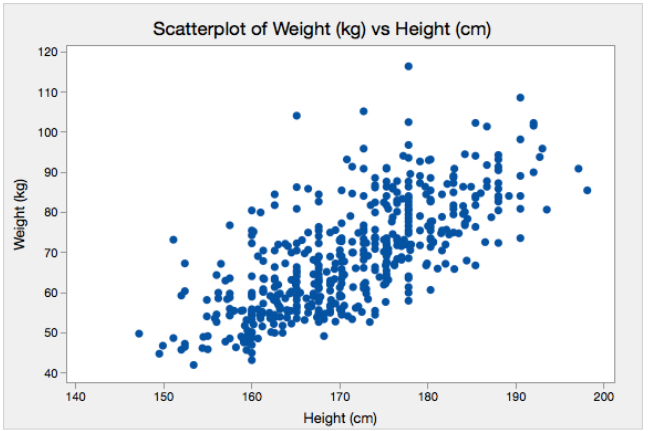
- Graphs
- Charts
- Maps
- Diagrams

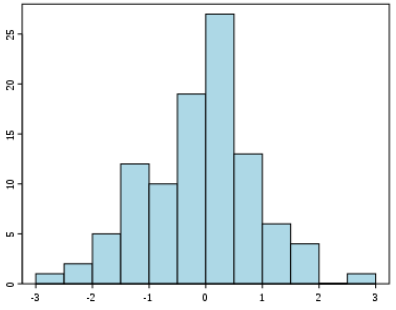
S. Kosslyn, “Understanding charts and graphs”, Applied Cognitive Psychology, 1989.

Graphs

- Visual display that illustrates one or more relationships among entities
- At least two scales are required
- Shorthand way to present information
- Allows a trend, pattern or comparison to be easily comprehended

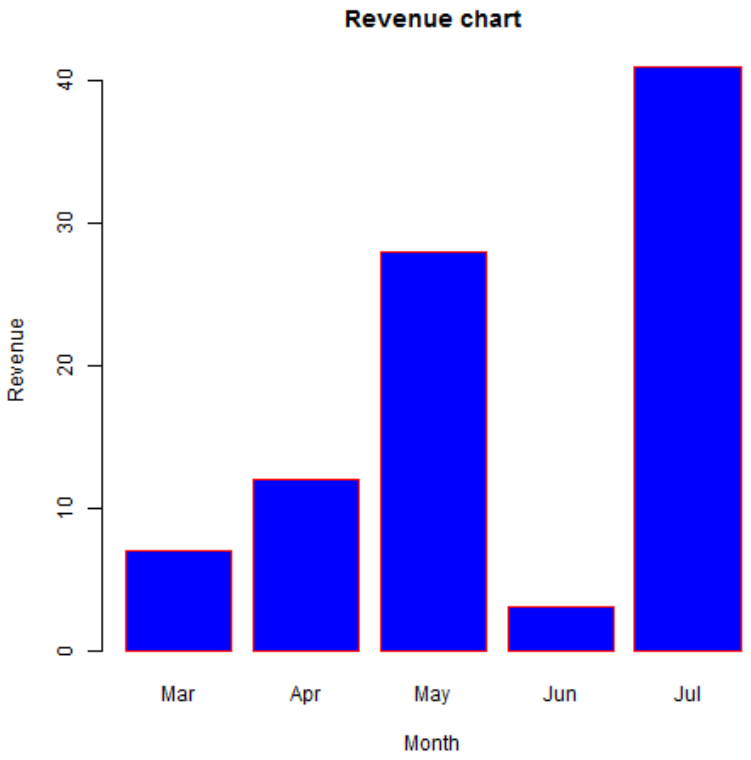






Bar graph

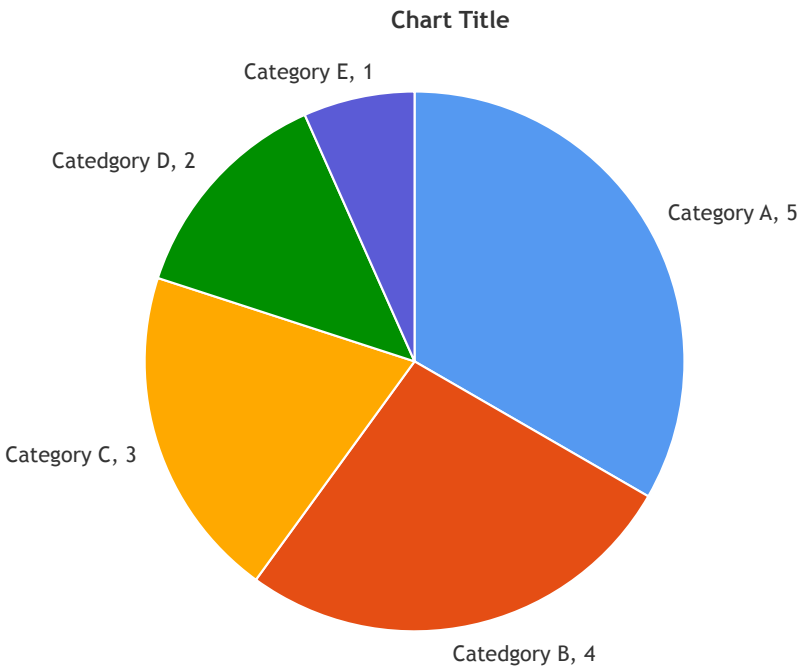
- Graphical display of data using bar
- It can be horizontal or vertical
- Bars provide a visual presentation of categorical data
- The bar height are scaled to their values
- Good for comparisons



Bar graphs are good for ranking relations, in which quantitative values are sequenced by size, from large to small or vice versa. It is also sufficient for nominal comparison

Pie graph

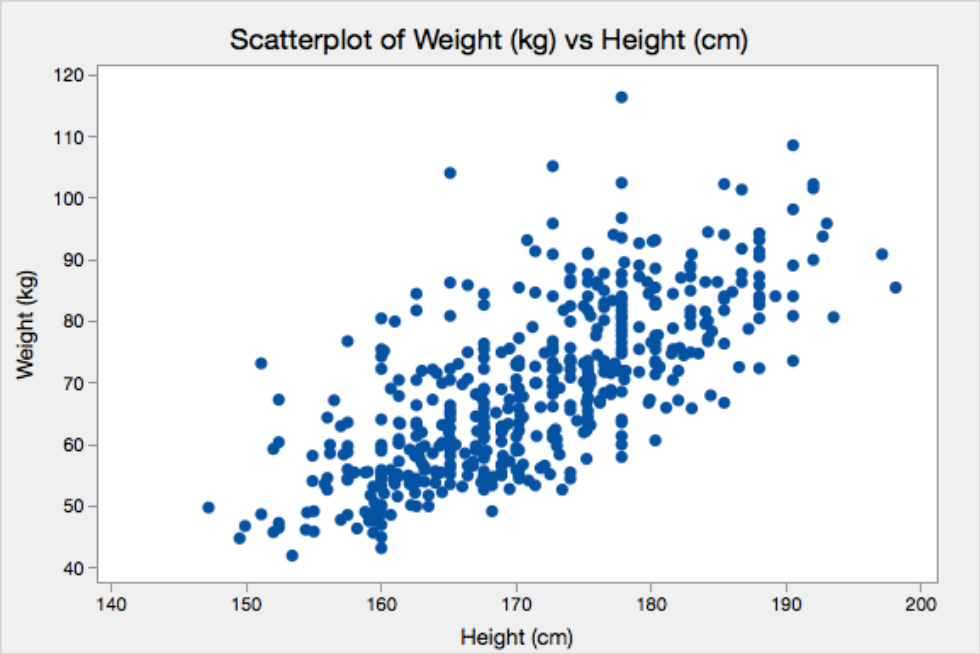
- A type of graph in which a circle is divided into slices
- Slices show relative size of data



It is good for plotting part- to-whole relations, in which quantitative values are displayed to reveal the portion that each value represents to some whole, this is called a

Scatterplot

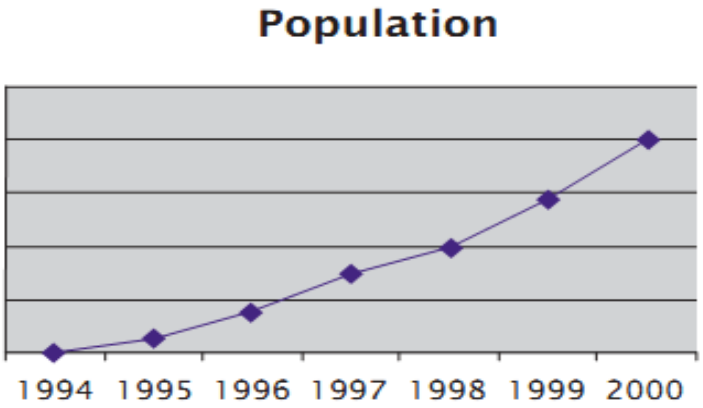
- A scatter has points that show the relationship between two sets of data
- It displays paired data using y and x axes



- It is good to show distribution of data, how a set of quantitative values are spread across their entire range
- It is also good to show correlation between two variables.

Line graph

- Normally used to show a change
- Gridlines can help the eye to navigate the plot
- If comparing 2 graphs use the same scale
- Use a sensible number of tick points and labeling
- Make comparing two attributes easy
- ALWAYS label the axis

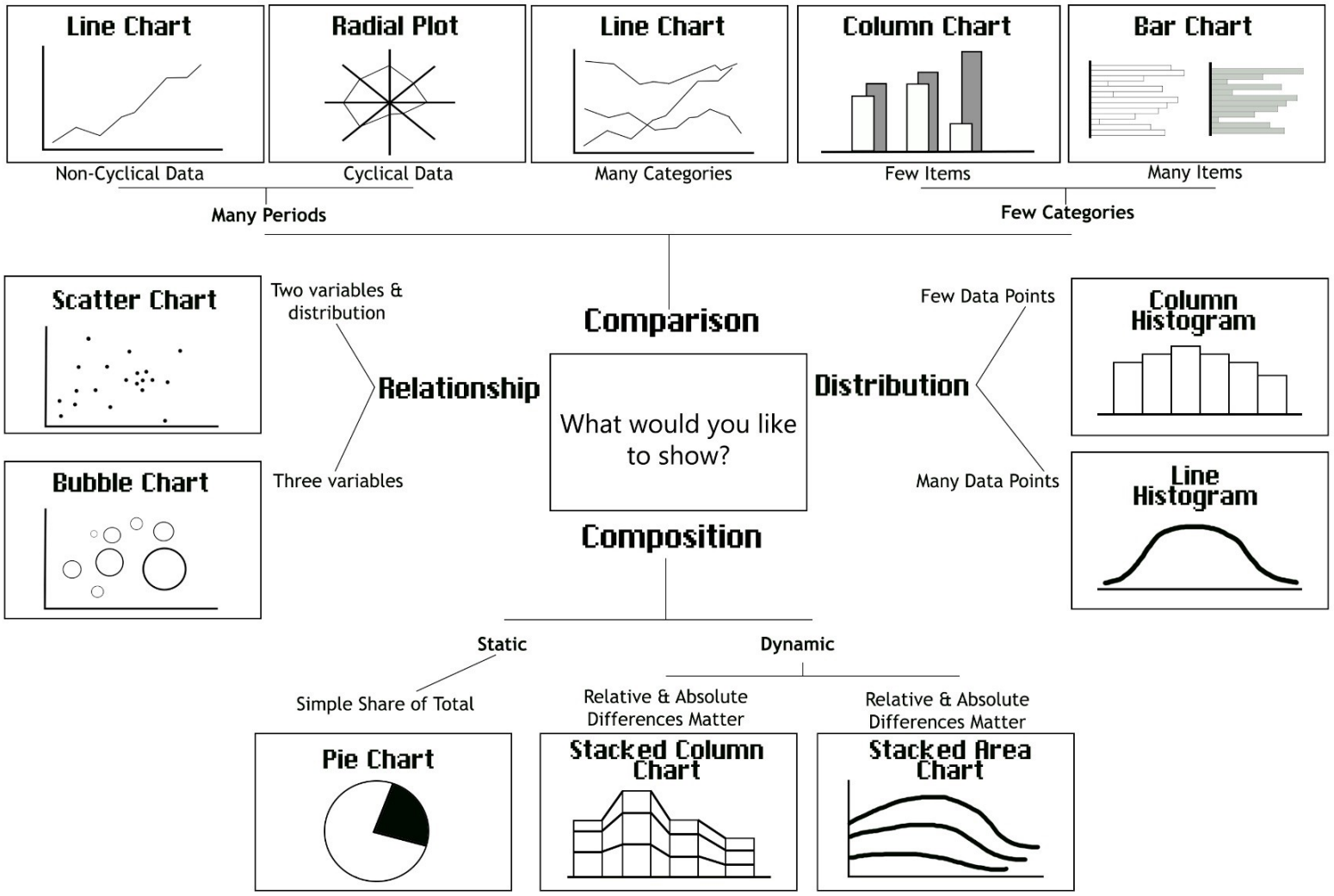


- It is good for plotting time series relations, in which quantitative values are expressed as a series of measures taken across equal intervals of time
- Time series reveal trends and patterns

Other graphs

- http://www.datavizcatalogue.com

Graphs usage



http://guides.auraria.edu/c.php?g=528873&p=3876041

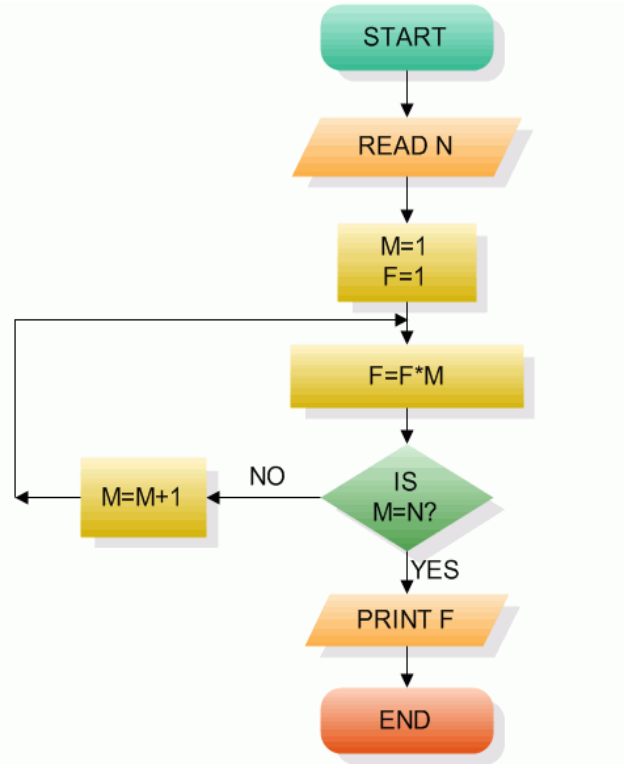
Basic symbolic displays

- Graphs
- Charts
- Maps
- Diagrams

S. Kosslyn, “Understanding charts and graphs”, Applied Cognitive Psychology, 1989.

Charts

- Structure is important, relates entities to each other
- Primarily uses lines, enclosure, position to link entities
- Discrete relations among discrete entities
- Structure relates entities to one another
- Example: flowchart, network diagram



Basic symbolic displays

- Graphs
- Charts
- Maps
- Diagrams

S. Kosslyn, “Understanding charts and graphs”, Applied Cognitive Psychology, 1989.

Maps

- Representation of spatial relations
- Locations identified by labels

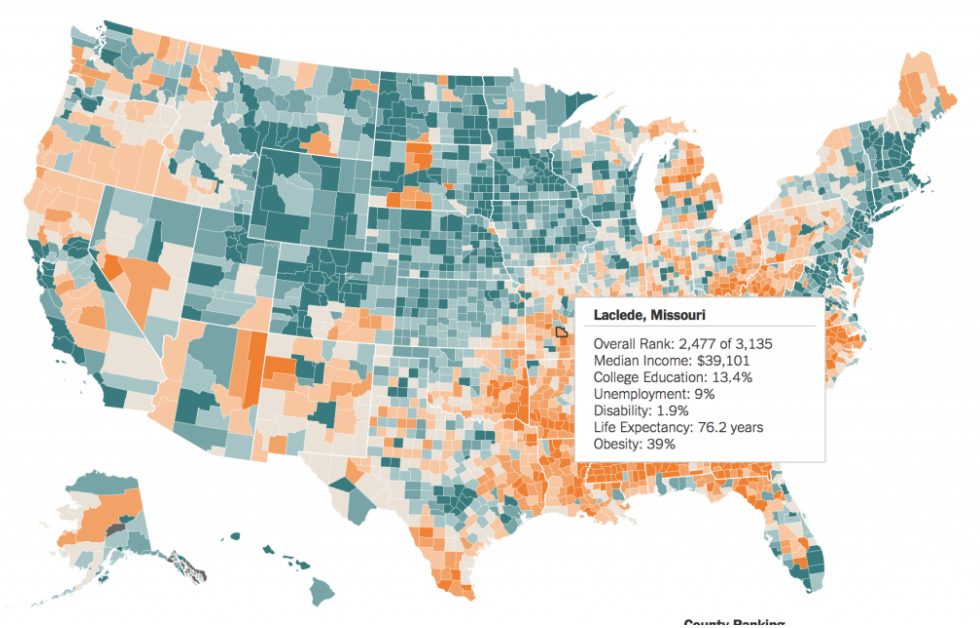


Image source: https://www.maptive.com/7-interactive-maps-that-prove-data-visualization-is-not-boring-2/

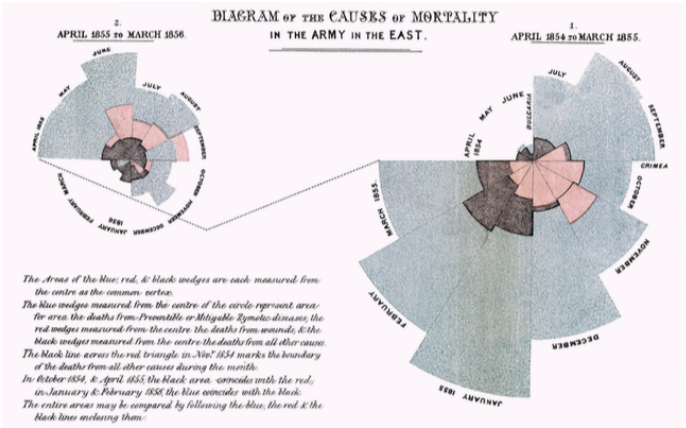
Basic symbolic displays

- Graphs
- Charts
- Maps
- Diagrams

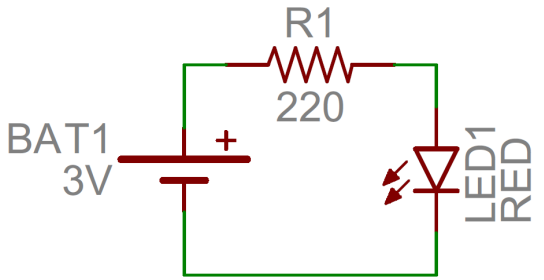
S. Kosslyn, “Understanding charts and graphs”, Applied Cognitive Psychology, 1989.

Diagrams

- Schematic picture of object or entity
- Parts are symbolic
- Examples: figures, steps in a manual, how-to illustrations



http://en.wikipedia.org/wiki/File:Nightingale-mortality.jpg



https://cdn.sparkfun.com/assets/d/2/9/c/c/51cdbd5fce395fc80b000000.png

Course evaluation

- Complete this survey

Data analysis and visualization in R

Links for code
notebook format, pdf format

Data visualization using Tableau

Download data from here

Project

- Milestone 1 is due on Sept. 29th
  - Submit one page that contains
    - the project title
    - a listing of project members
    - the topic to be addressed
    - potential data sources
- 1 min elevator pitches on Oct. 3th
  - Only one member of the team presents

Project title
Title
Date: 2017/09/26

Group Members

- Member's name
- Member's name
- Member's name
- Member's name

Description

Write project description here.

Potential Dataset

Provide a link or name data source(s) here.

Assignment 3
Due Oct. 10th

Next Lecture

- Intro to text analysis and visualization