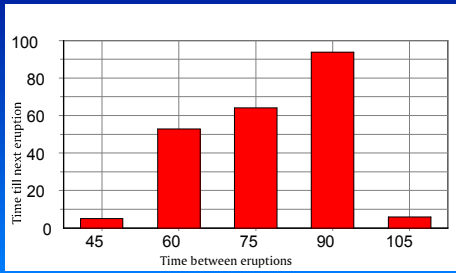


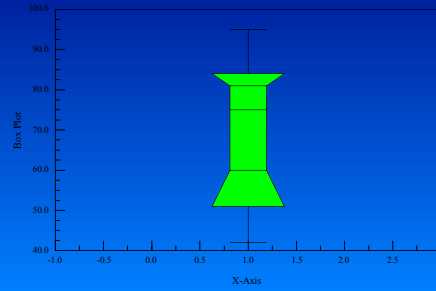
Categorical Representations

Histogram of Old Faithful Eruptions



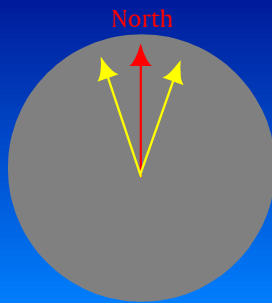
Same Data, Different Representation

GEYSERI.DAT



What is the average of 15 and 345?

- 180 lbs
- North



Types of Data

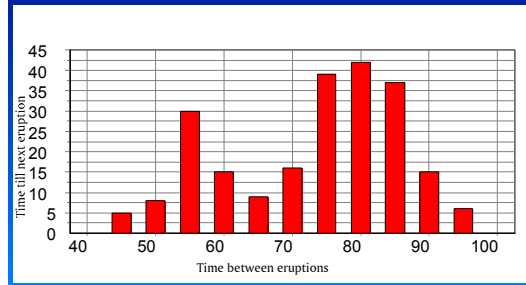
- Ratio: (CEO income)/(avg worker income)
- Interval: temperature (boiling, freezing)
- Closed: concentration (must add to 100%)
- Directional: $0^\circ = 360^\circ$
- Ordinal: ranking (Moh's hardness)
- Discrete: countable (number of organisms)
- Nominal or categorical: list (names)
- Continuous: infinitely subdivided (weight)

Converting between types

Reducing the number of possible values

- Intervals: reduce continuous data to discrete or counts to frequency
- Frequency tables: reduce non-continuous data to intervals by counting data
 - Relative frequency: conversion of frequency to percentage of total counts
 - Histogram: bar chart of frequency table
- Categories: chosen to simplify representation (*birds* instead of *crows*, *pigeons*, etc.)

Effect of choosing intervals



The Trade-off

- Reduce complexity of representation
 - Allows patterns to emerge
 - Easier to recognize major trends
- Lose detail
 - The anomalies may matter
 - Deviations from what we expect may be the key to future research

Mississippi River Discharge

Baton Rouge, LA

- Open with Excel the txt file from email
 - Specify that it is delimited by spaces
 - Locate the year, month, and discharge columns
 - Add a new column next to month and create a yearly fraction in decimal form
 - Plot date versus discharge
- Copy data, sort by month, and calculate monthly average discharge
 - Plot month versus average discharge
 - Is there a pattern?

In class

- Chapter 6 exercises
