Statistics One

Lecture 2

Descriptive Statistics

Three segments

- Histograms
- Summary statistics
- Tools for inferential statistics

Lecture 2
Segment 1

Histograms

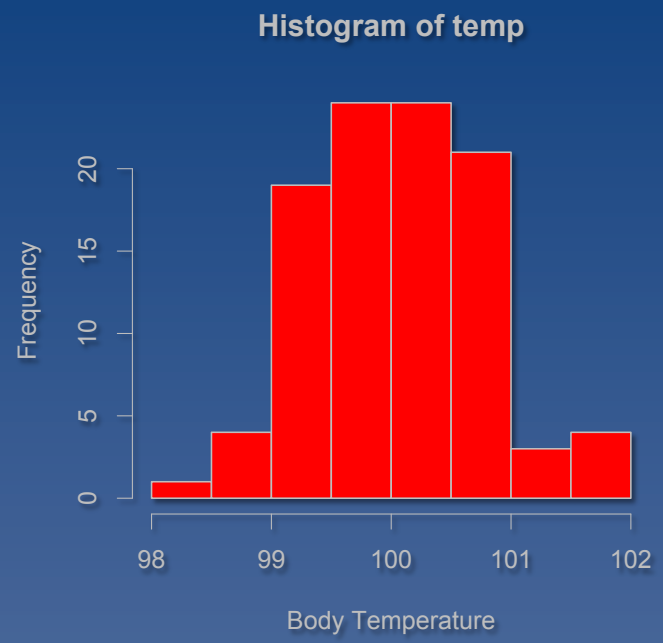
Histograms

- Important concepts
  - What is a distribution?
  - The normal distribution
  - Many non-normal distributions
    - Skew
    - Kurtosis

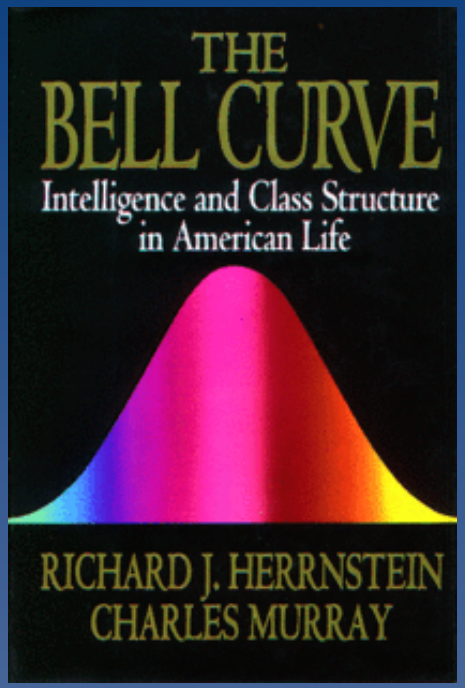
Histograms

- First, why start with histograms?
  - To overcome the natural tendency to rely upon summary information, such as an average
  - Histograms show an entire distribution

An example: Body temperature



A well-known histogram



Histograms

- Often reveal important information
  - Sex differences in spatial reasoning
  - In some studies, the mean for males is higher than the mean for females BUT
  - The variance within each group is far greater than the gap between the groups

Histograms

- To observe the sex difference:
  - Independent variable
    - Sex (female, male)
  - Dependent variable
    - Score on a test of spatial reasoning
      - For this example, let's assume a scale of 0-100
  - Concerns?

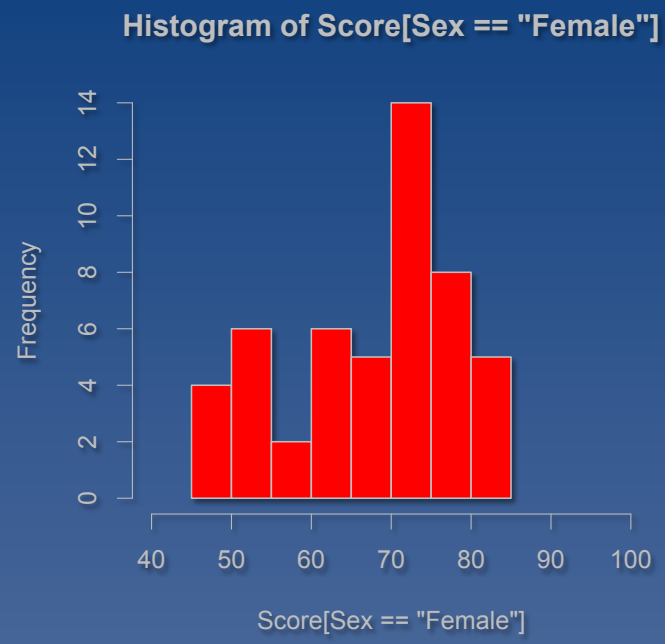
Histograms

- Simulated data
  - 50 females
  - 50 males
- Mean for females = 67.92
- Mean for males = 72.62

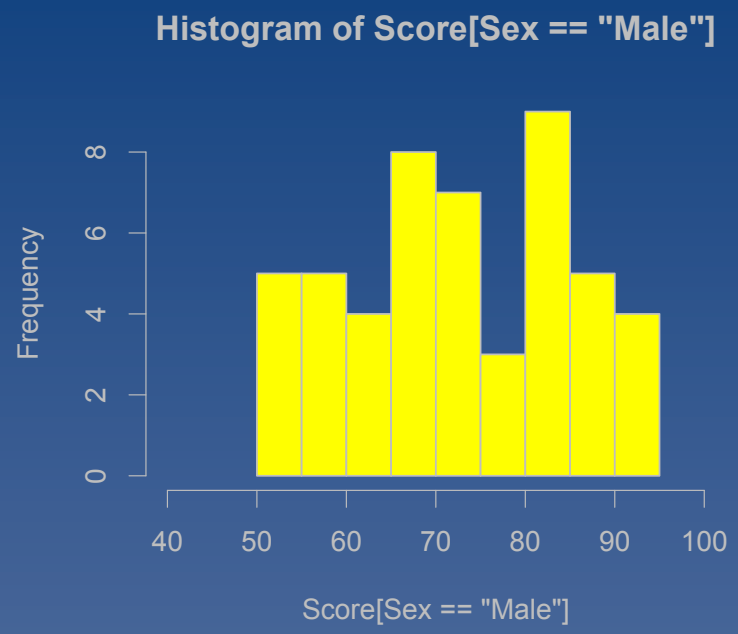
Histograms

- BUT compare the histograms
  - Females in red
  - Males in yellow

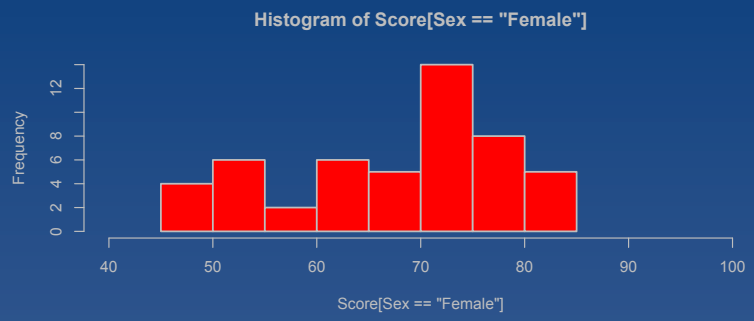
Histogram of females

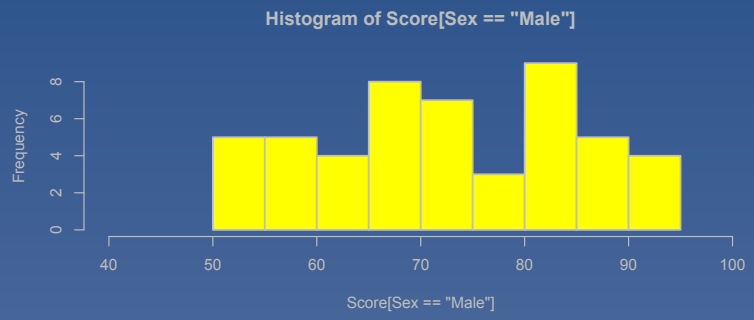


Histogram of males



Comparison of histograms





Histograms

- Another example
  - Does contraception have an effect on blood pressure?
  - In the 1960s when women started taking "the pill" there was some concern that the pill caused an increase in blood pressure

Histograms

- A drug study was conducted
  - Independent variable
    - Contraception (users, non-users)
  - Dependent variable
    - Blood pressure
      - For this example, we're plotting systolic blood pressure
  - Concerns?

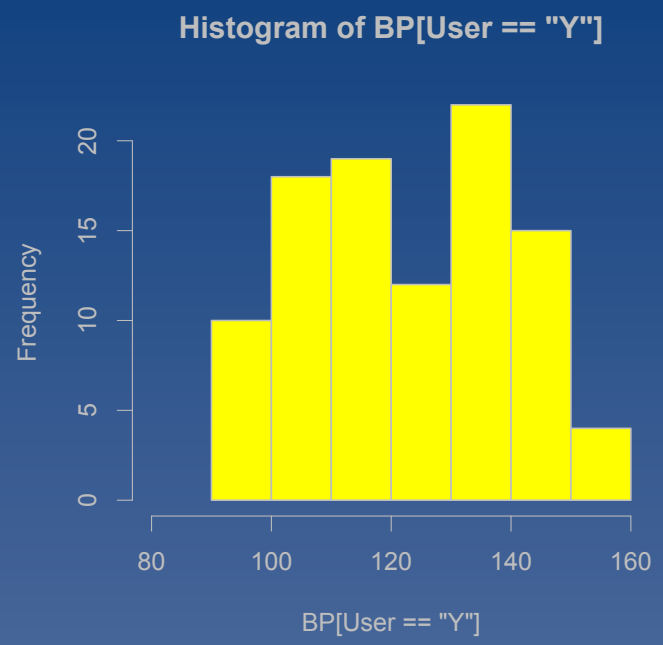
Histograms

- Simulated data
  - 100 users
  - 100 non-users
- Mean for users = 124.41
- Mean for non-users = 119.85

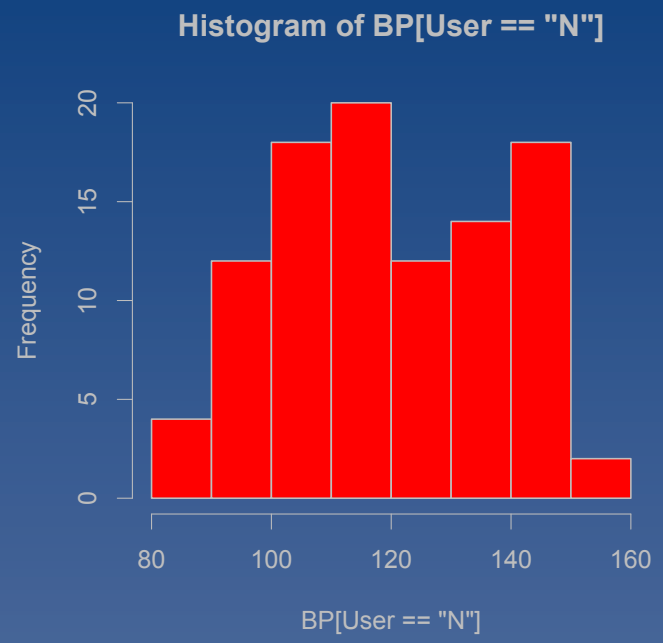
Histograms

- Let's look at the histograms
  - Users in yellow
  - Non-users in red

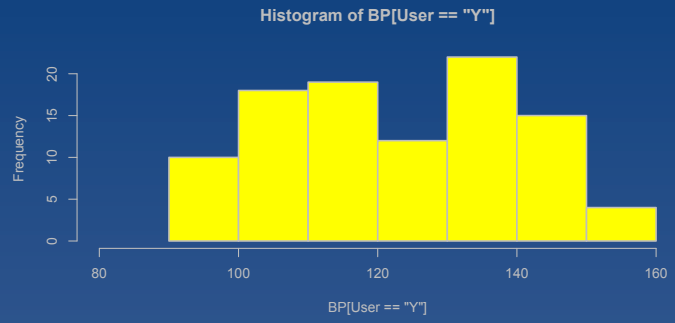
Histogram of users

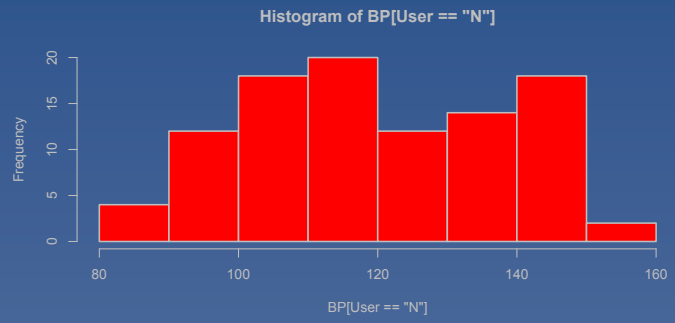


Histogram of non-users



Comparison of histograms





Non-normal distributions

- A few examples
  - Uniform
  - Positively skewed
  - Negatively skewed

Wine tasting!

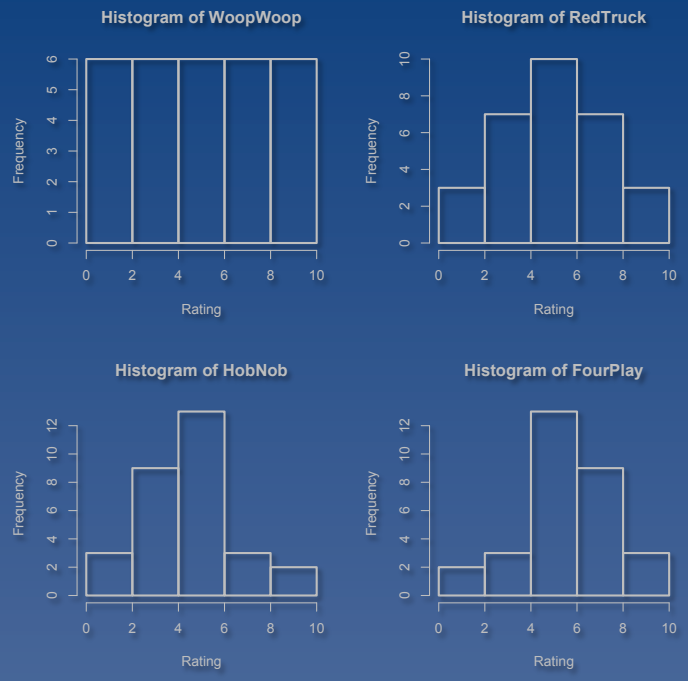
Example

- Suppose that 30 wine experts rated the overall quality of 4 different wines on a scale of 1-10
  - Higher scores indicate higher quality

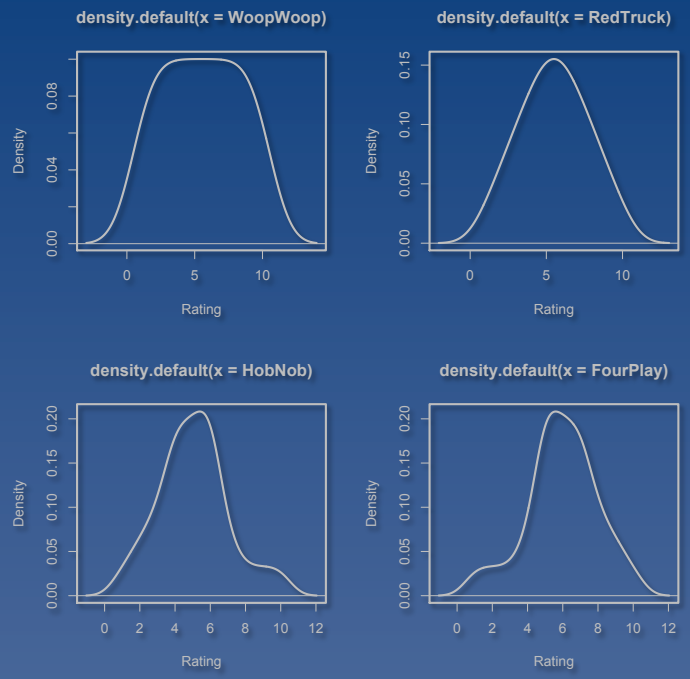
Example

- Four red wines
  - WoopWoop
  - RedTruck
  - HobNob
  - FourPlay

Four histograms



Four density plots



Summarize a distribution

- Central tendency
- Variability
- Skew
- Kurtosis

  – In mathematics, these are known as the four "moments" of the mean

Histograms: Review

- Important concepts
  - What is a distribution?
  - The normal distribution
  - Many non-normal distributions
    - Skew
    - Kurtosis

Image in slide 7 was retrieved from
http://en.wikipedia.org/wiki/File:TheBellCurve.gif.

Image in slide 23 was retrieved from
http://www.delawareonline.com/blogs/secondhelpings/
uploaded_images/redwinegl-758416.JPG