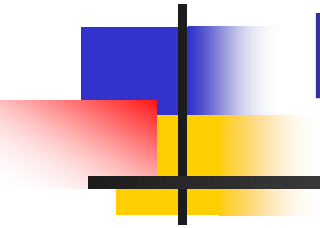
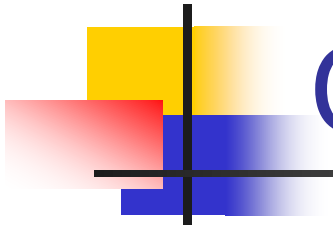


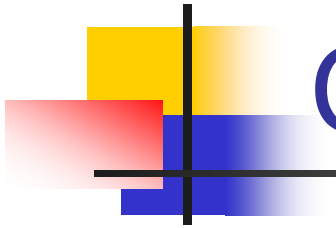
Module 1 – Section 1

Numerical and Graphical Summaries of a Categorical Variable



Outline

- Categorical Variables
- Numerical Summaries
- Graphical Summaries



Categorical Variables

- Variable whose values belong to one or more categories
- Two Types
  - Nominal - categories do not have a natural order
  - Ordinal - categories have a natural order

Examples: Nominal Categorical Variables

- Eye Color
  - Blue
  - Brown
  - Green
  - Hazel
  - Other
- Political Party
  - Democrat
  - Republican
  - Independent

Examples: Ordinal Categorical Variables

- Grade in High School
  - Freshman
  - Sophomore
  - Junior
  - Senior
- Attitudes about Premarital Sex
  - Always Wrong
  - Almost Always Wrong
  - Sometimes Wrong
  - Not Wrong At All

Ordinal Categorical Variables

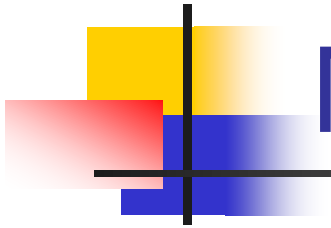
- Ordinal data with set distance between categories are sometimes treated as quantitative, not categorical.
  - Ex. Likert-scale surveys – 1 to 5, 1 to 6, etc.
- Must assume distance between categories is the same.

Ex. Variables

- Students enrolled in STAT 101 over the course of several semesters responded to a survey. Two of the questions on the survey were:
  - What is your gender?
    - Options given: Male, Female
  - What is your eye color?
    - Options given: Blue, Brown, Green, Hazel, Other

Ex. Data – Gender and Eye Color

Gender	Eye Color
Female	Blue
Male	Blue
Male	Blue
Female	Blue
⋮	⋮
⋮	⋮
Male	Other
Female	Other



Numerical Summaries

- Frequency Table
- Relative Frequency Table
- Summary Table

Frequency Table

- Gives number or count in each category
- Nominal Categories can be presented in any order.
- Ordinal Categories are presented in their natural order.
  - Most statistical packages use alphabetical order by default – user can change order.

Ex. Frequency Table

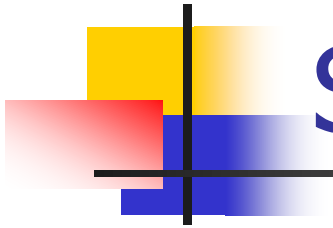
Gender	Freq
Female	1125
Male	943
Total	2068

Relative Frequency Table

- Gives proportion or percentage in each category
  - Round proportions to 4 to 6 decimal places
  - Round percentages to 2 to 4 decimal places
- Nominal Categories can be presented in any order
- Ordinal Categories are presented in their natural order

Ex. Relative Frequency Table

Eye Color	Prop
Blue	0.35251
Hazel	0.16780
Green	0.14894
Brown	0.31044
Other	0.02031
Total	1.00000



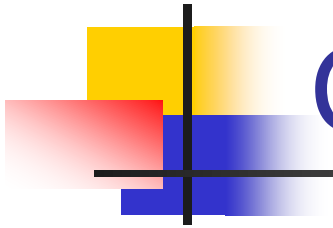
Summary Table

- Typical tabular display of a categorical variable
- Combination of frequency and relative frequency tables.

Ex. Summary Tables

Gender	Freq	Prop
Female	1125	0.5440
Male	943	0.4560
Total	2068	1.0000

Eye Color	Freq	Prop
Blue	729	0.35251
Hazel	347	0.16780
Green	308	0.14894
Brown	642	0.31044
Other	42	0.02031
Total	2068	1.00000



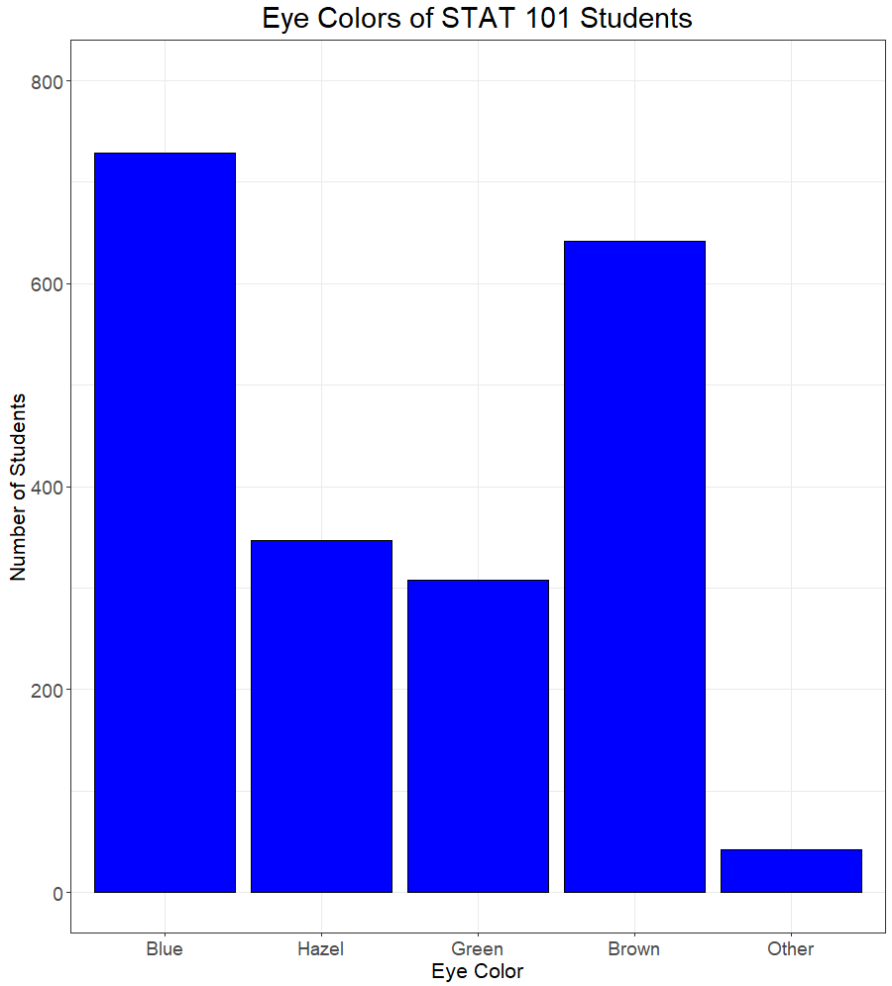
Graphical Summaries

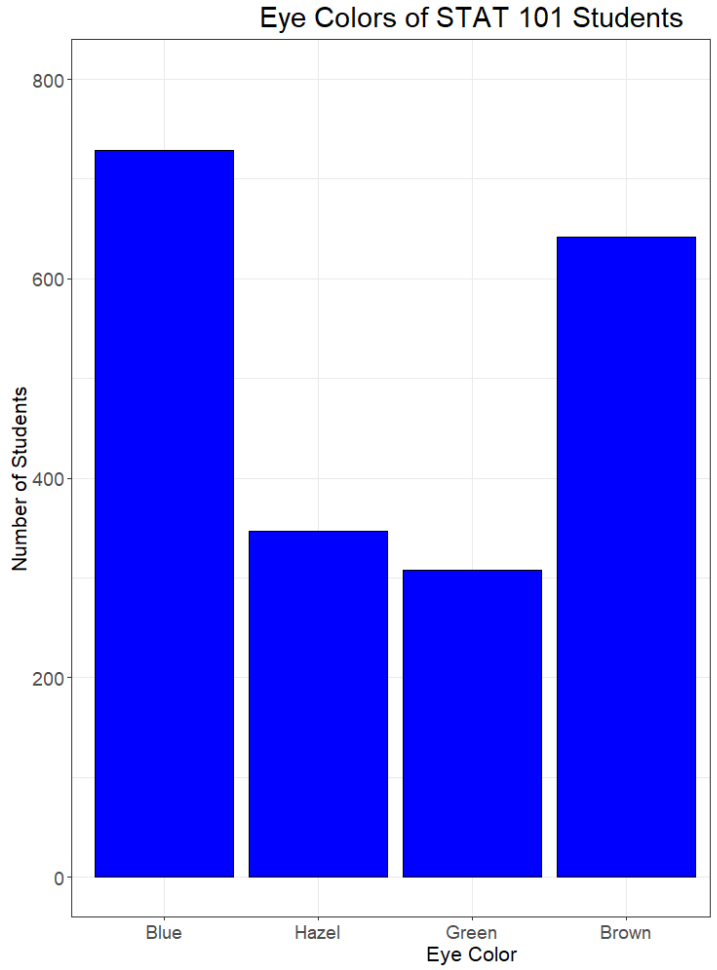
- Bar Graph
- Pie Chart

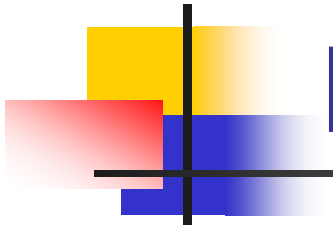
Bar Graph

- Displays either **number** or percentage for each category
- Numbers or percentages = heights of bars
- Can leave out categories if desired

Ex. Bar Graph


Eye Colors of STAT 101 Students
800
600
400
200
0
Number of Students
Blue
Hazel
Green
Brown
Other
Eye Color

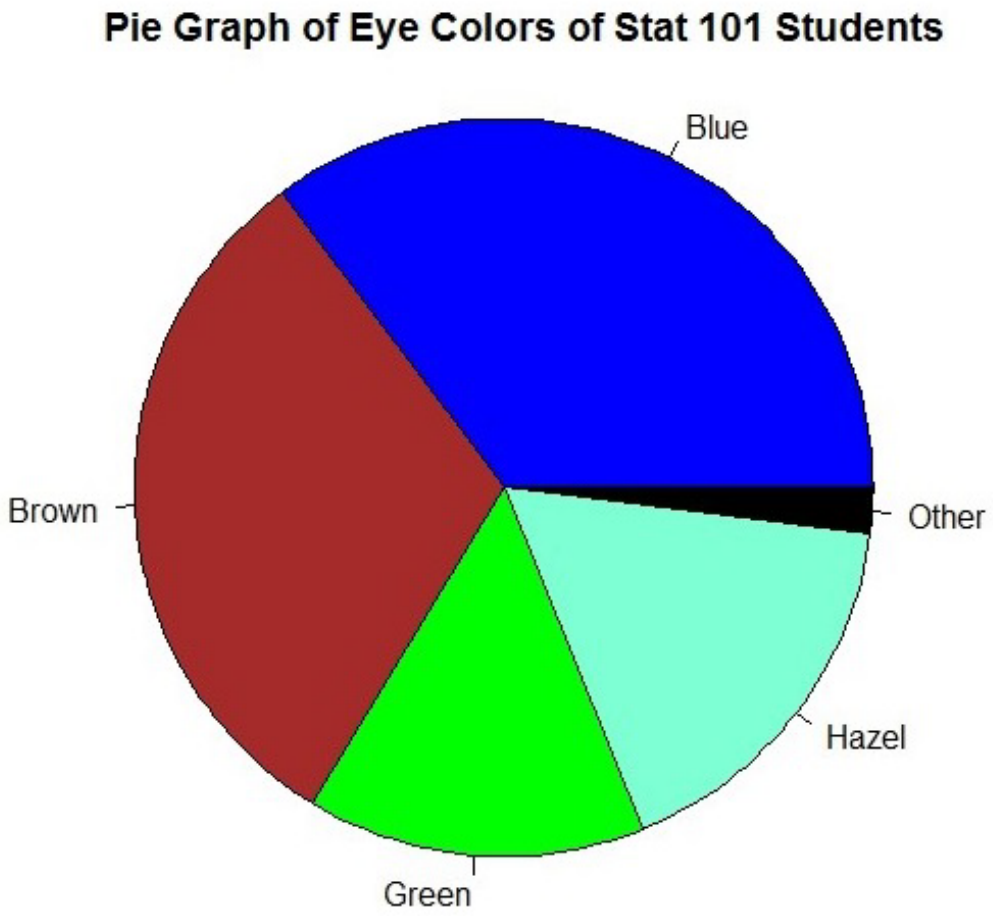

Eye Colors of STAT 101 Students
800
600
400
200
0
Number of Students
Blue
Hazel
Green
Brown
Eye Color

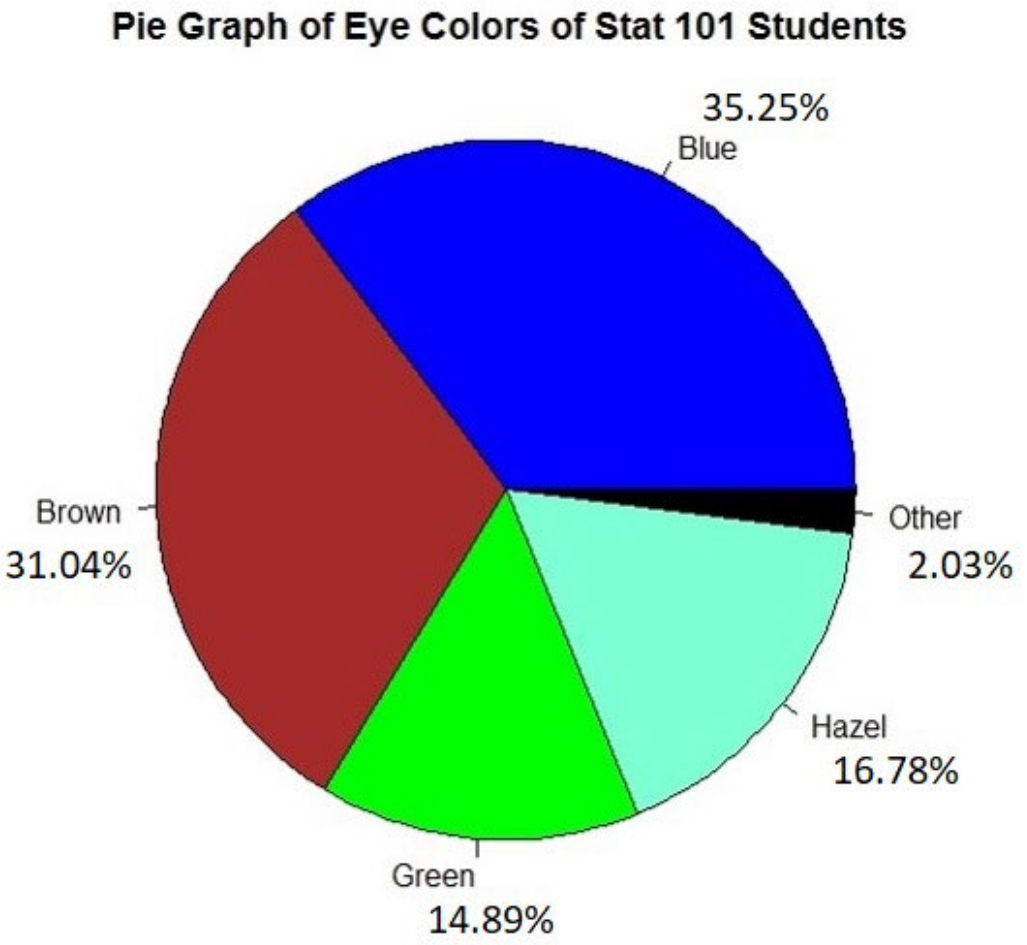


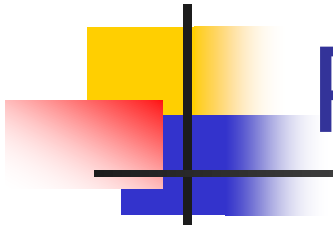
Pie Chart

- Displays percentage of whole for each category
- Compare sizes of pie slices
- Must have all categories in display
  - Percentage of whole
- Should include percentages in display

Ex. Pie Graph

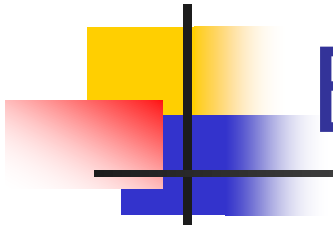

Pie Graph of Eye Colors of Stat 101 Students
Blue
Other
Brown
Hazel
Green


Pie Graph of Eye Colors of Stat 101 Students
35.25%
Blue
Brown
31.04%
Other
2.03%
Hazel
16.78%
Green
14.89%



Pie Charts

- Appear in publications more often
- Difficult to compare pie slices without percentages
- Difficult to extend to more than one variable



Bar Graphs

- Visually easier for us to compare numbers or percentages for categories.
- Easier to extend to more than one variable
- Are confused with histograms – summary of distribution of a single quantitative variable