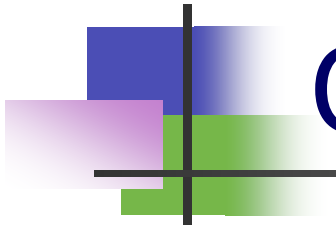
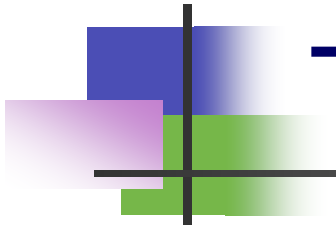
Module 2 – Section 1

Numerical and Graphical Summaries
of a Contingency Table



Outline

- Variables, Data, and the Contingency Table
- Marginal Distributions
- Conditional Distributions
- Graphical Summaries



Two Categorical Variables

- Variable 1
  - Categorical with I categories.
- Variable 2
  - Categorical with J categories.

Ex. Data

- 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	Eye Color
Female	Blue
Male	Blue
Male	Blue
Female	Blue
⋮	⋮
⋮	⋮
Male	Other
Female	Other

Ex. Variables from STAT 101 Students

- Variable 1 – Gender
  - Categories (Female, Male)
- Variable 2 – Eye Color
  - Categories (Blue, Hazel, Green, Brown, Other)

Contingency Table

- Cross-classify observations according to categories on both variables.
- Y_{ij} = number of observations in the i^{th} category of variable 1 **and** j^{th} category of variable 2.

Contingency Table (3 x 4)

	Variable 2				
Variable 1	Cat 1	Cat 2	Cat 3	Cat 4	Total
Cat 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Cat 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Cat 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	n

Ex. Contingency Table – Eye Color and Gender

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729	347	308	642	42	2068

Marginal Distributions

- Distribution of one categorical variable, ignoring the other.
  - Variable 1 = $Y_{i.}$ for $i = 1, \ldots, I$
  - Variable 2 = $Y_{.j}$ for $j = 1, \ldots, J$

Marginal Distribution: Variable 1

	Variable 2				
Variable 1	Cat 1	Cat 2	Cat 3	Cat 4	**Total**
Cat 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$\boldsymbol{Y_{1.}}$
Cat 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$\boldsymbol{Y_{2.}}$
Cat 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$\boldsymbol{Y_{3.}}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	$\boldsymbol{n}$

Ex. Marginal Distribution – Gender

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729	347	308	642	42	2068

Ex. Marginal Distribution – Gender (with %)

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125 (54.40)
Male	359	160	110	290	24	943 (45.60)
Total	729	347	308	642	42	2068 (100)

Marginal Distribution: Variable 2

	Variable 2				
Variable 1	**Cat 1**	**Cat 2**	**Cat 3**	**Cat 4**	Total
Cat 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Cat 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Cat 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Total	$\boldsymbol{Y_{.1}}$	$\boldsymbol{Y_{.2}}$	$\boldsymbol{Y_{.3}}$	$\boldsymbol{Y_{.4}}$	$\boldsymbol{n}$

Ex. Marginal Distribution – Eye Color

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729	347	308	642	42	2068

Ex. Marginal Distribution – Eye Color (with %)

	Eye Color					
Sex	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729 (35.25)	347 (16.78)	308 (14.89)	642 (31.04)	42 (2.03)	2068 (100)

Conditional Distributions

- Distribution of Variable 2 given category of Variable 1
  - $Y_{ij}, j = 1, \ldots, J$ for fixed i
  - I conditional distributions

Conditional Distributions

	Variable 2				
Variable 1	Cat 1	Cat 2	Cat 3	Cat 4	Total
Cat 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Cat 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Cat 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	n

Ex. Conditional Distributions – Eye Color given Gender

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729	347	308	642	42	2068

Ex. Conditional Distributions – Eye Color Given Gender (with %)

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370 (32.89)	187 (16.62)	198 (17.60)	352 (31.29)	18 (1.60)	1125 (100)
Male	359 (38.07)	160 (16.97)	110 (11.66)	290 (30.75)	24 (2.55)	943 (100)
Total	729	347	308	642	42	2068

Conditional Distributions

- Distribution of Variable 1 given category of Variable 2
  - $Y_{ij}, i = 1, \ldots, I$ for fixed j
  - J conditional distributions

Conditional Distributions

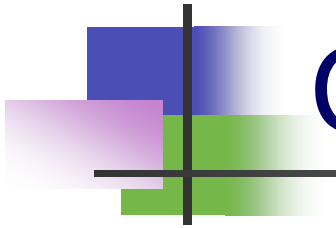
	Variable 2				
Variable 1	Cat 1	Cat 2	Cat 3	Cat 4	Total
Cat 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Cat 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Cat 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	n

Ex. Conditional Distributions of Gender given Eye Color

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370	187	198	352	18	1125
Male	359	160	110	290	24	943
Total	729	347	308	642	42	2068

Ex. Conditional Distributions – Gender given Eye Color (with %)

	Eye Color					
Gender	Blue	Hazel	Green	Brown	Other	Total
Female	370 (50.75)	187 (53.89)	198 (64.29)	352 (54.83)	18 (42.86)	1125
Male	359 (49.25)	160 (46.11)	110 (35.71)	290 (45.17)	24 (57.14)	943
Total	729 (100)	347 (100)	308 (100)	642 (100)	42 (100)	2068



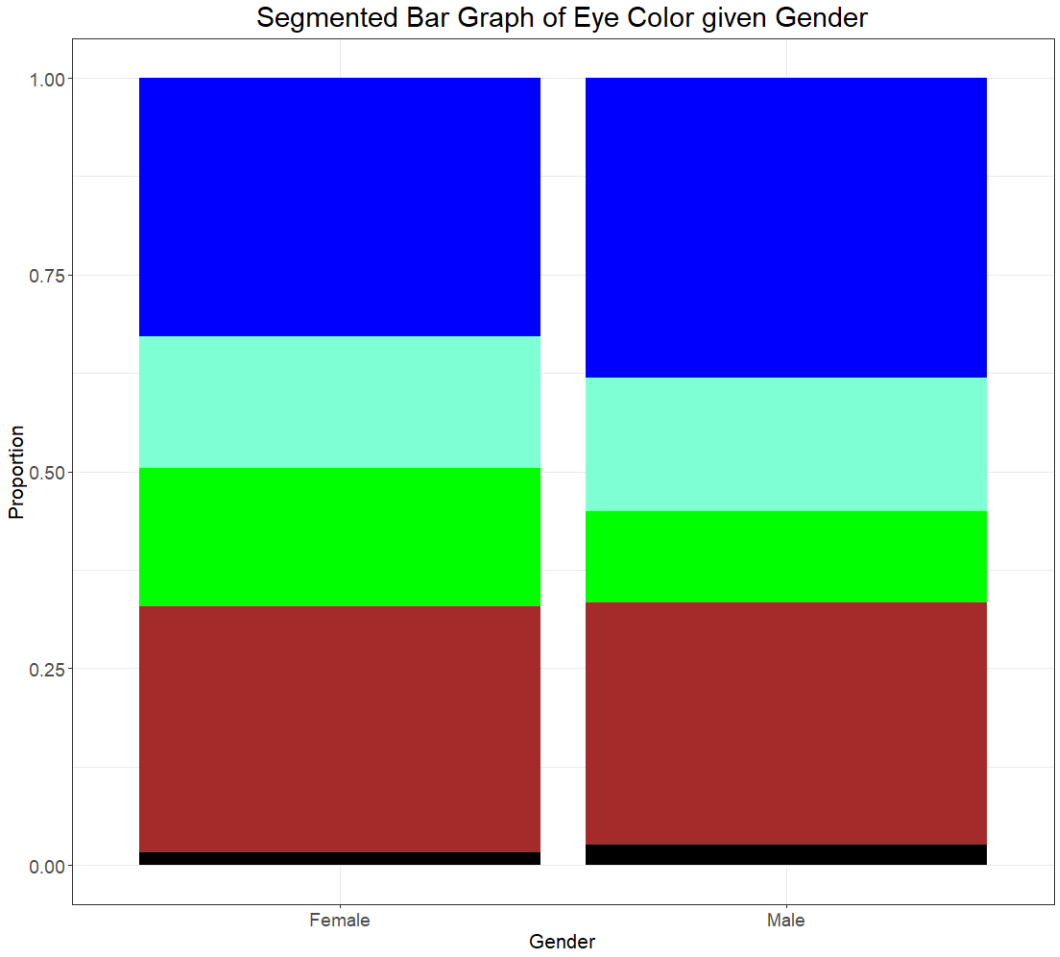
Graphical Summaries

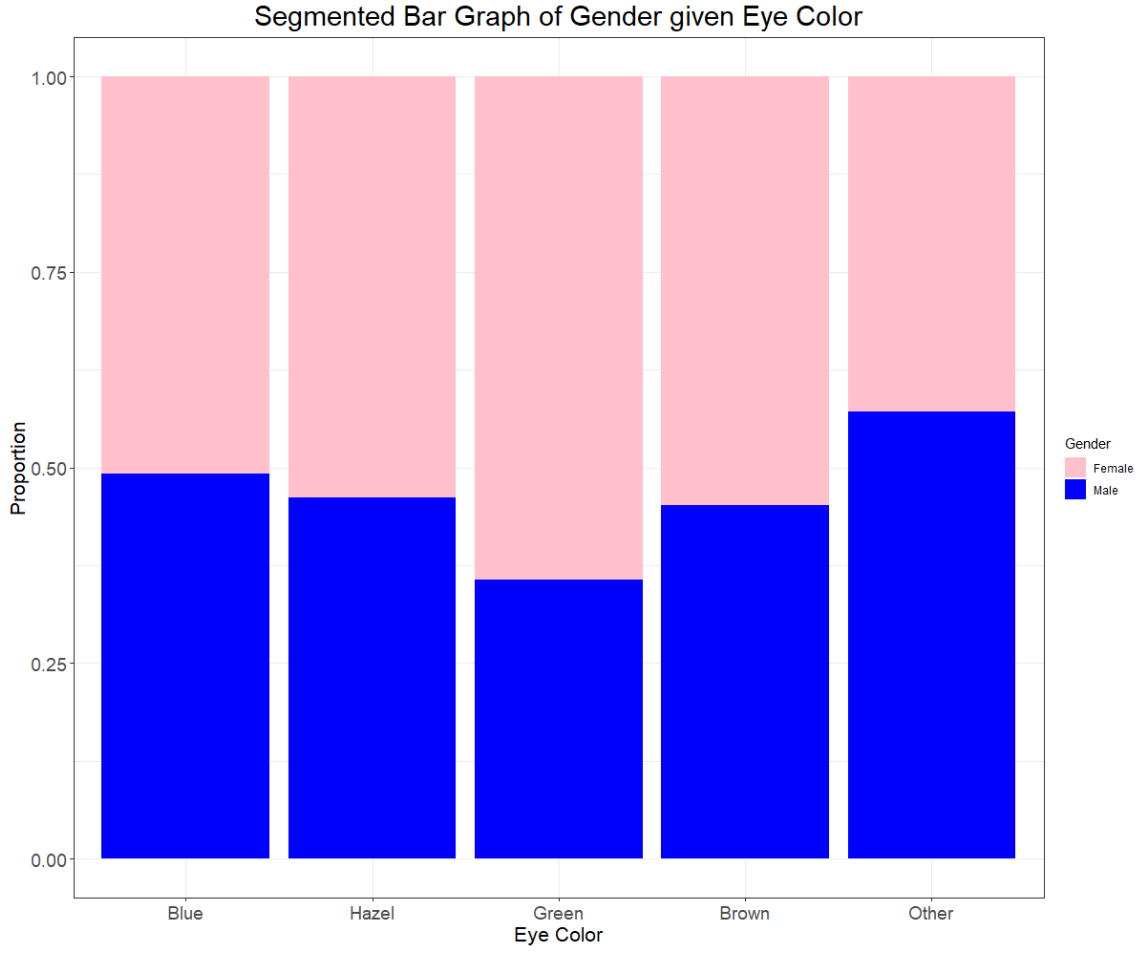
- Segmented Bar Graphs
- Mosaic Plot

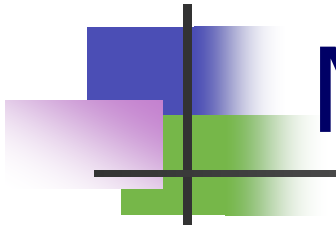
Segmented Bar Graph

- Stacked Bar Graphs of one set of Conditional Distributions
  - Ex. Eye Color Given Gender
  - Ex. Gender Given Eye Color
- Same Color for each category in one of the variables
- Each Stack is a category in the other variable

Ex. Eye Color and Gender







Mosaic Plot

- Similar to Segmented Bar Graph
- Width of each stack is based on the variable's marginal distribution

Ex. Eye Color and Gender