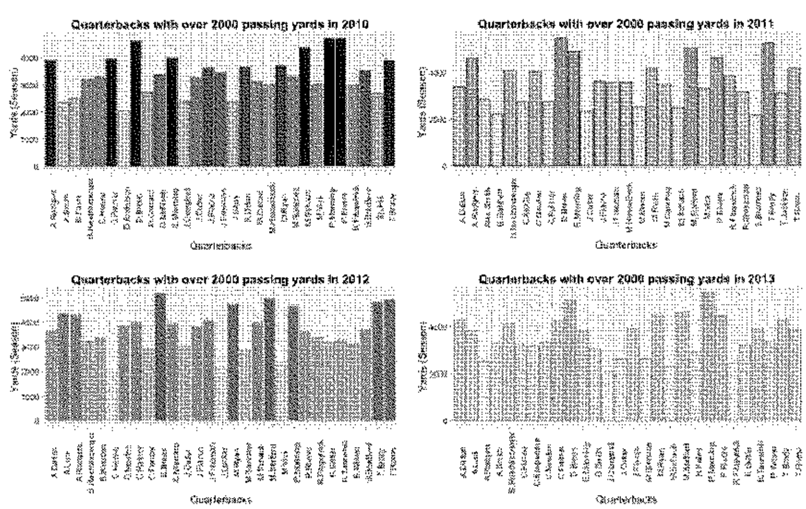
Today: Marginal and Conditional Distributions
2-D Caterogical Data
Contingency Tables and Mosaic Plots
Friday: 2-D Categorical

Sam Ventura
36-315

Department of Statistics
Carnegie Mellon University

September 14, 2016

Quarterbacks with Over 2000 Yards, 2010 – 2013



Stacked Bar Chart

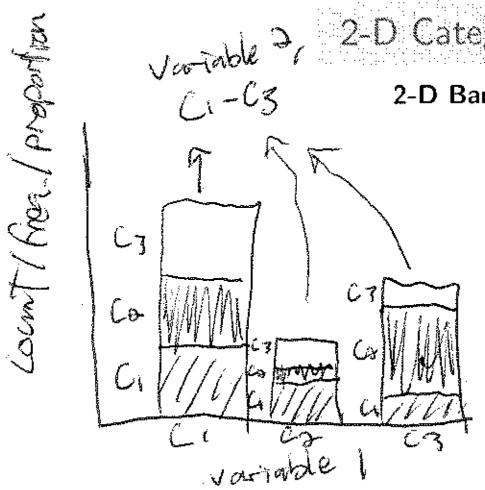
"Bar chart of spine charts"

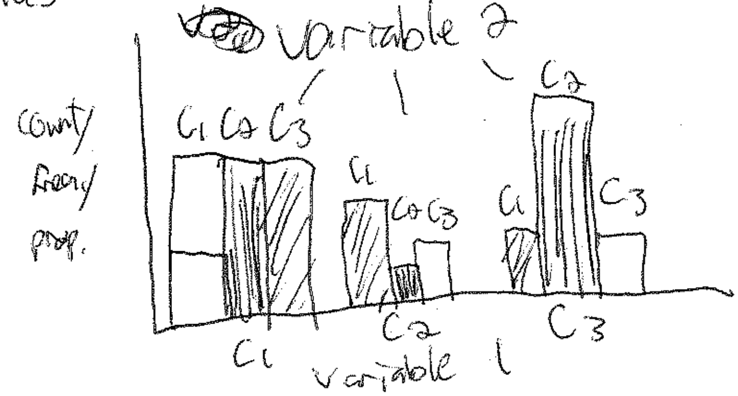
Side-by-Side Bar Chart

"Bar chart of bar charts"

2-D Categorical: Stacked and Side-by-Side Bar Charts

2-D Bar Charts: Two choices





What does a 2-D bar chart show?:

easier to make comparisons of variable 1 categories

easier to make comparisons of variable 2 ~~categories~~ given variable 1

- Marginal dist'n of V1
- several conditional distributions of Variable 2 given each category of Variable 1

-empirical conditional dist'n

3/8

Contingency Tables

Contingency Tables Give Us:

Empirical
- Marginal Distribution (of variable dist'n one)
- Joint Distribution (of 2 Variables dist'n simultaneously ~~and variable 2~~)
- Conditional Distributions (dist'n of one variable given a particular value of another variable)

Two Categorical Variables:

set up:

Variable 1: K_1 categories ; Variable 2: K_2 categories

What Are We Interested In?

- observed counts in each cell
  - ↳ O_{ij} = # of obs w/ var1 $= C_1$ & Var2 $= C_2$
- Expected counts in each cell
  - ↳ $E_{ij} = \dfrac{n_{i\cdot} \times n_{\cdot j}}{n}$

4/8

n = total # of obs

$n_{i\cdot}$ = total in row i

$n_{\cdot j}$ = total in col j

→ all of the O_{ij}'s

Joint = entire table (minus margins)

Contingency Tables and Marginal/Conditional Distributions

Var 2

		C_1^*	C_2^*	$\cdots$	$C_{K_2}^*$	
Var 1	C_1	O_{11}	O_{12}	$\cdots$	O_{1K_2}	$n_{1\cdot}$
	C_2	O_{21}	O_{22}	$\cdots$	O_{2K_2}	$n_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	C_{K_1}	$O_{K_1 1}$	$O_{K_1 2}$	$\cdots$	$O_{K_1 K_2}$	$n_{K_1\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot K_2}$	n

→ Marginal distn of Var 1

$\boxed{\text{each row}}$ = conditional distribution of variable 2 given a category of Var 1

$\boxed{\text{each column}}$ = cond. distn of Var 1 given a cat. of Var 2

→ marginal distn of Var 2

Recall: Independence Rules from Probability

Can input contingency tables into chi-square tests for independence

E.g. `chisq.test(table(var1, var2))`

More on this in Lab 04