

Today: 2-D Categorical Data Independence and Mosaic Plots

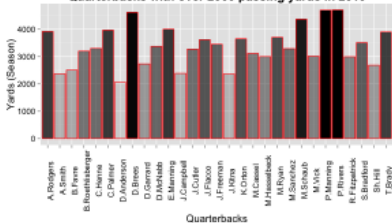
Sam Ventura
36-315

Department of Statistics
Carnegie Mellon University

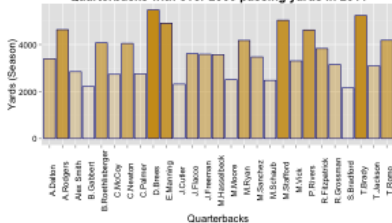
February 1, 2016

Quarterbacks with Over 2000 Yards, 2010 – 2013

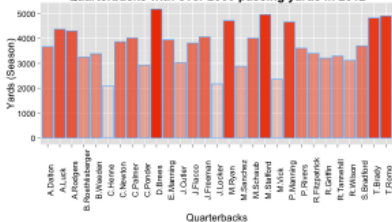
Quarterbacks with over 2000 passing yards in 2010



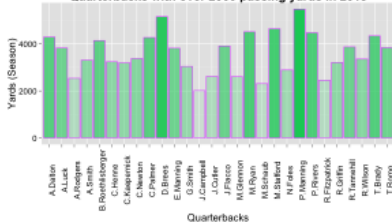
Quarterbacks with over 2000 passing yards in 2011



Quarterbacks with over 2000 passing yards in 2012



Quarterbacks with over 2000 passing yards in 2013



Last Time: 2-D Bar Charts

Stacked Bar Charts:

Side-By-Side Bar Charts:

Contingency Tables

Two Categorical Variables:

Asympototics:

Contingency Tables and Marginal/Conditional Distributions

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

Pearson Residuals

Pearson Residuals: Scaled difference between observed/expected

Mosaic Plots

Mosaic Plots: Area plot for two categorical variables

Can color the boxes by their differences from what was expected:

Wednesday: Mosaic Plot Demo