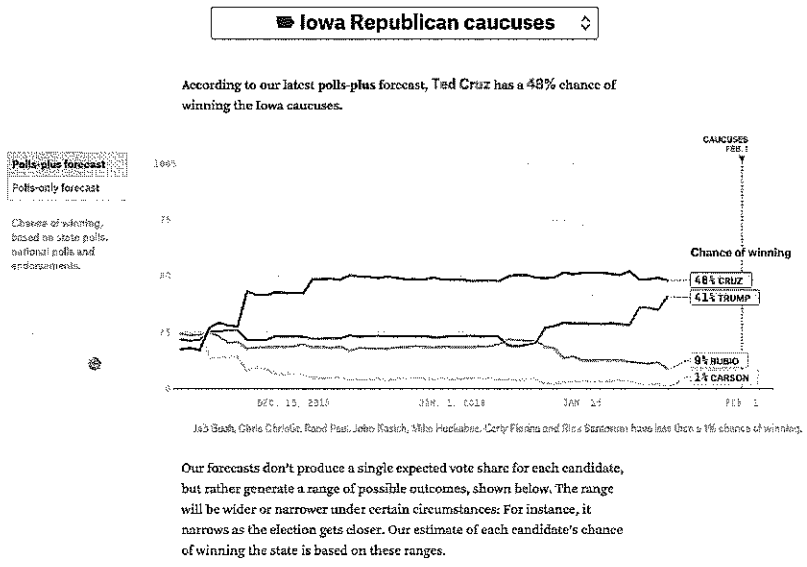
Today: 1-D Categorical
Bar, Spine, Rose, Pie
Statistical Tests for Categorical Data
Wednesday: 2-D Categorical

Sam Ventura
36-315

Department of Statistics
Carnegie Mellon University

January 25, 2016

Via 538: Chances of Winning Iowa Republican Caucuses



Via 538: Chances of Winning in New Hampshire (Democrats)

Chance of winning New Hampshire

	POLLS-ONLY FORECAST	POLLS-PLUS FORECAST
Sanders	86% ▲	61% ▲
Clinton	14% ▼	39% ▼
O'Malley	<1%	<1%

or proportions

Chi-Square Test for Independence

Chi-squared test: Statistical test used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories (of a categorical variable).

2-D Categorical: Used to test differences in the conditional distributions (more on this Wednesday)

1-D Categorical:

K = # of unique categories

p_K = proportion of obs. in category K

set-up #1

H_0: $p_1 = p_2 = p_3 = p_4 = \dots = p_K$

H_a: at least one is different

In R, chisq.test()

set-up #2

H_0: $p_1 = p_1^*, p_2 = p_2^*, \dots, p_K = p_K^*$

H_a: at least one is not the same as expected.

p_K^* = expected proportion in cat. K