Week 3 | Charting data

Jue Yang, technologist-in-residence

Ask me anything!
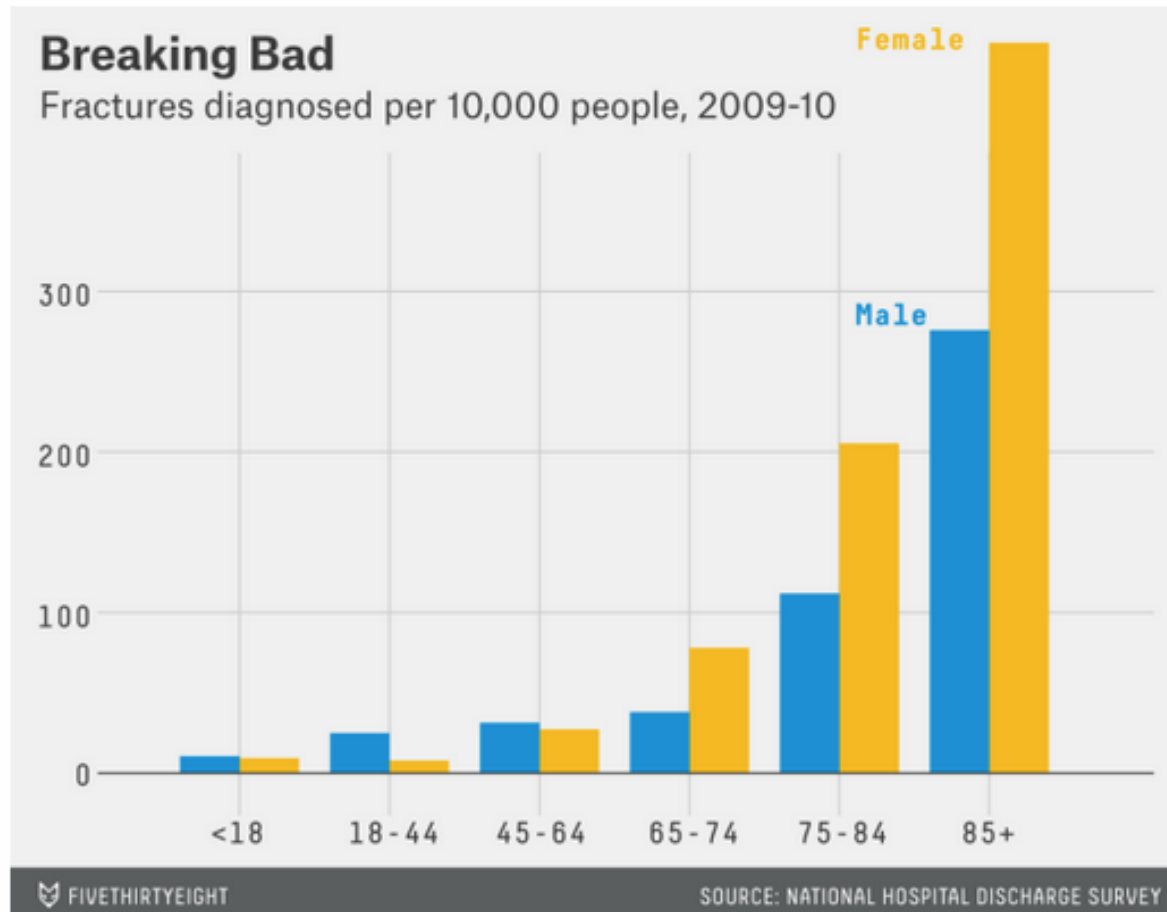
1. https://github.com/jueyang/call-me-maybe

2. https://chatwithjue.slack.com/

Basic charts types

- Bar
- Line
- Pie

Bar chart

Breaking Bad
Fractures diagnosed per 10,000 people, 2009-10

Female
Male

300
200
100
0

<18   18-44   45-64   65-74   75-84   85+

FIVETHIRTYEIGHT   SOURCE: NATIONAL HOSPITAL DISCHARGE SURVEY

Source: FiveThirtyEight

US Open

First grand slam to offer equal
prize money — 1973

Women earn less than men
Women achieve pay parity with men for singles title

100%
75
50
25
0

1970   1975   1980   1985   1990   1995   2000   2005   2010   2015
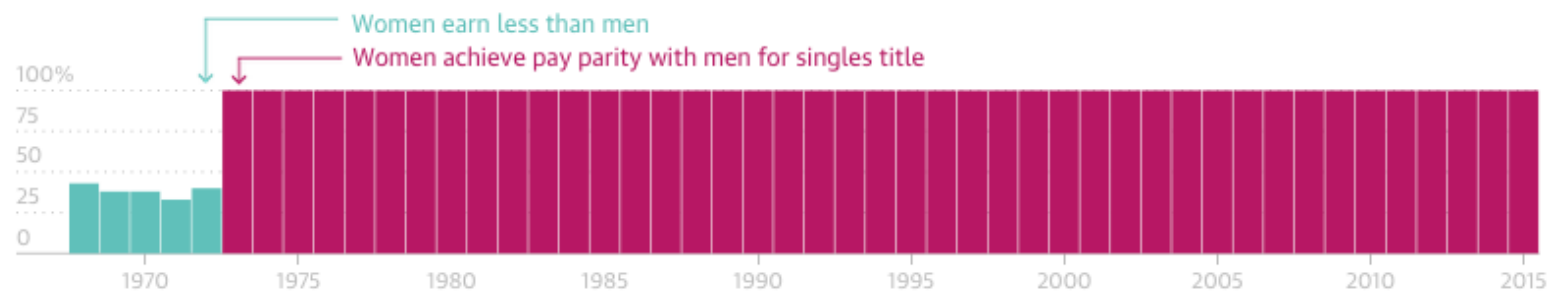
Source: The Guardian

Estimated share of NCAA athletes who will go pro

Sport	Share
Baseball	8.6%
Men's ice hockey	6.8
Football	1.6
Men's soccer	1.4
Men's basketball	1.2
Women's basketball	0.9
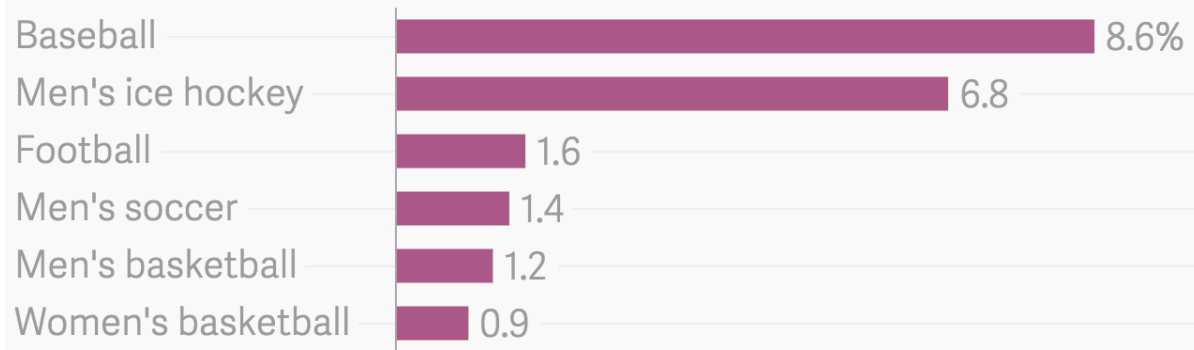
ATLAS | Data: NCAA

Source: Quartz
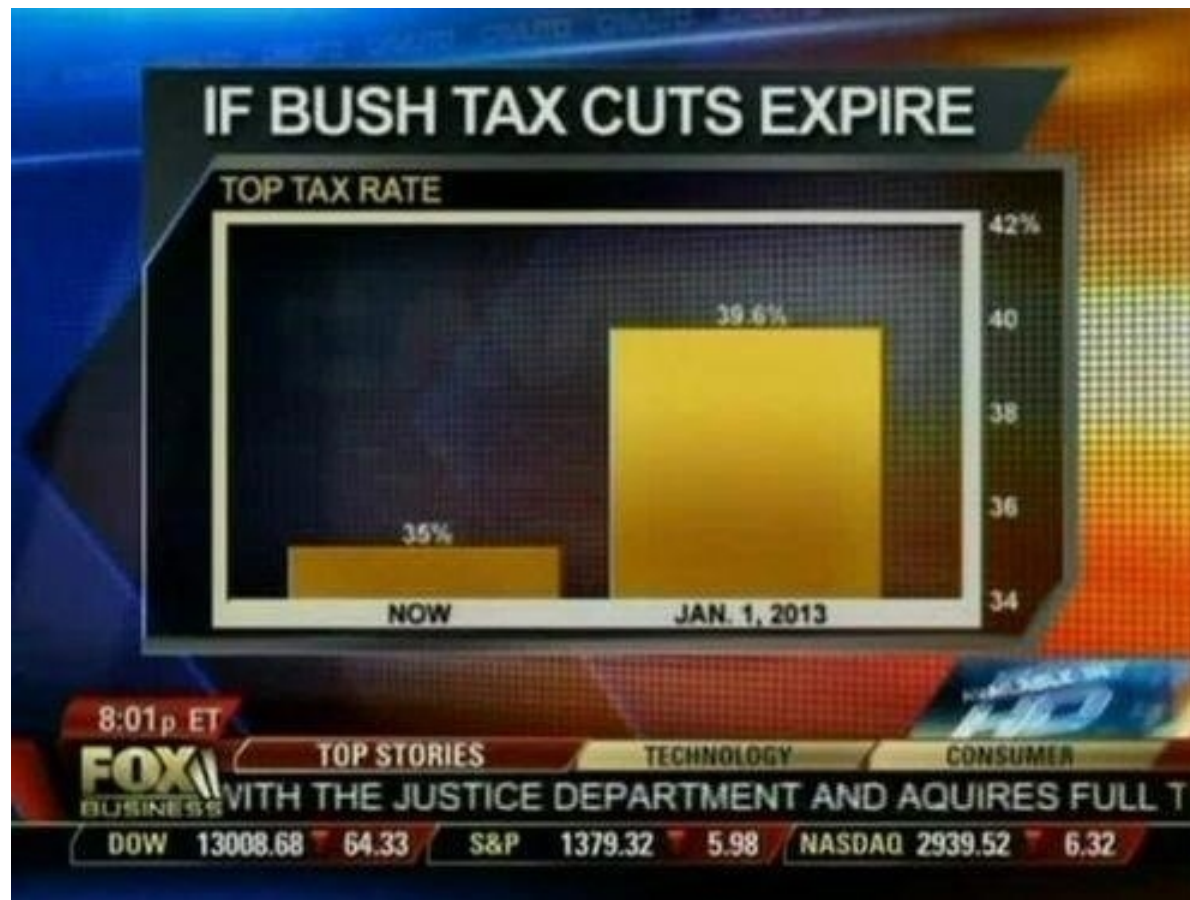
Source: Heap Analytics

Line chart

Brent crude oil price

$120 per barrel

100

80

60

40

Sep. Nov. '15 Mar. May Jul.

△TL△S | Data: FactSet
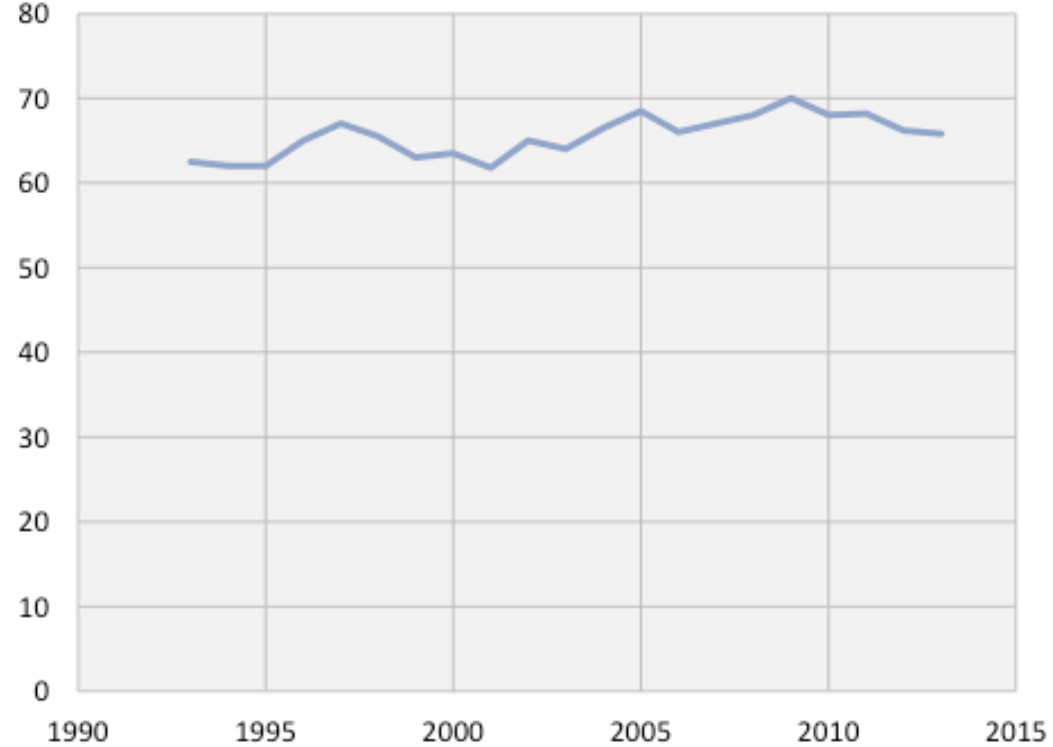
Source: Quartz
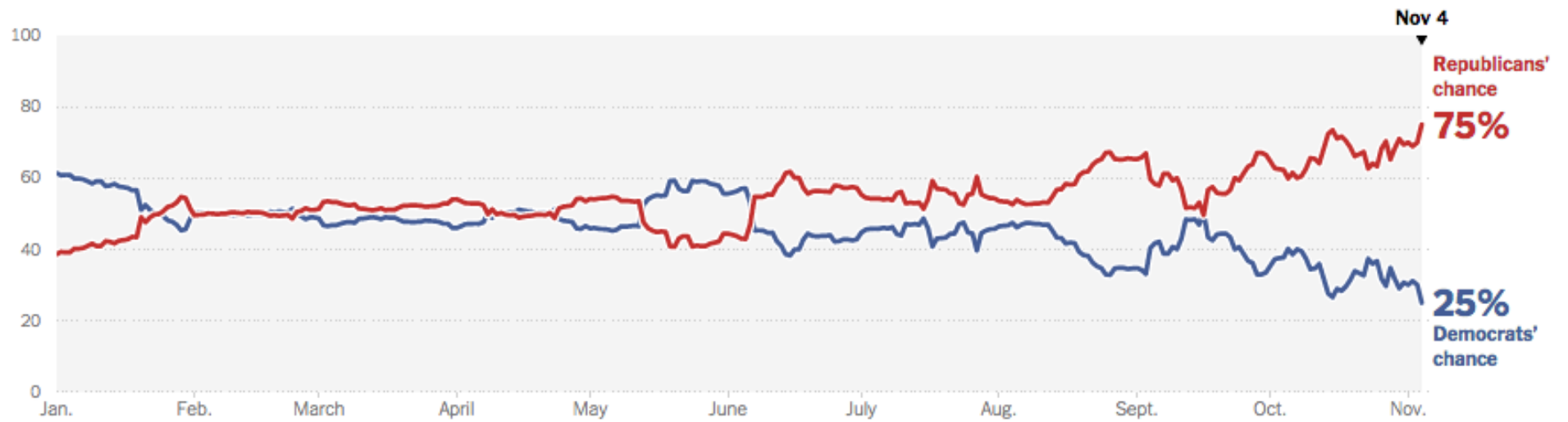
Source: New York Times

Enrollment Rate (%)

Source: Junk Charts

Pie chart

Detroit Property Taxes Due and Paid, 2011

31.7%

68.3%

Collected
Delinquent
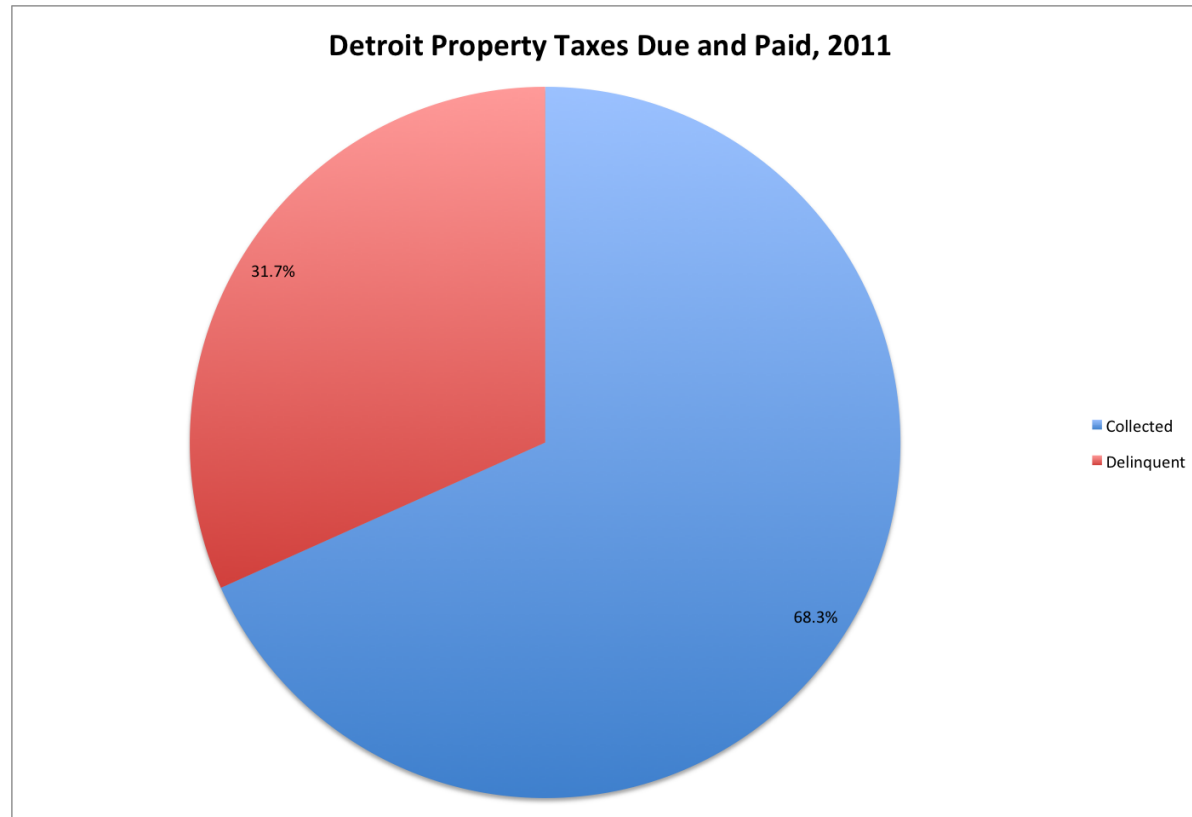
Source: Business Insider

Entire Stimulus Package, Billions of Dollars

state and local law enforcement, 4

???, 31.3

temporary increase in the Medicaid matching rate, 87

food stamps, 20

healthcare for newly unemployed, 39

increased unemployment benefits and job training, 43

preventative care, 4.1

health I.T., 20

higher education modernization, 6

Pell grant, 15.6

state fiscal relief, 79

local school districts, 41

transit and rail, 10

clean water, flood control, and environmental restoration investments, 19

tax cuts, 275

transform the nation's energy transmission, distribution, and production systems, 32

repair public housing and make key energy efficiency retrofits, 16

weatherize modest-income homes, 6

science facilities, research, and instrumentation, 10

expand broadband internet access, 6

highway construction, 30

modernize federal and other public infrastructure, 31

Source: House Committee on Appropriations

Source: New York Times

Time spent per activity

Hours per day by people aged 15-64, 1998-2009

Legend: Paid work and study | Unpaid work | Eating and sleeping | Personal care | Leisure | Other

France
4.2 / 3.2 / 2.0 / 2.2 / 1.3 / 11.1

Germany
3.9 / 3.2 / 3.1 / 3.1 / 0.8 / 10.0

Japan
6.3 / 2.5 / 2.4 / 2.2 / 0.9 / 9.8

Britain
4.4 / 3.1 / 2.7 / 3.3 / 0.8 / 9.8

United States
4.8 / 3.0 / 2.4 / 3.3 / 0.7 / 9.9

Turkey
4.0 / 3.4 / 2.1 / 3.3 / 0.7 / 10.5

Source: OECD

*Totals may not add up to 24 due to rounding
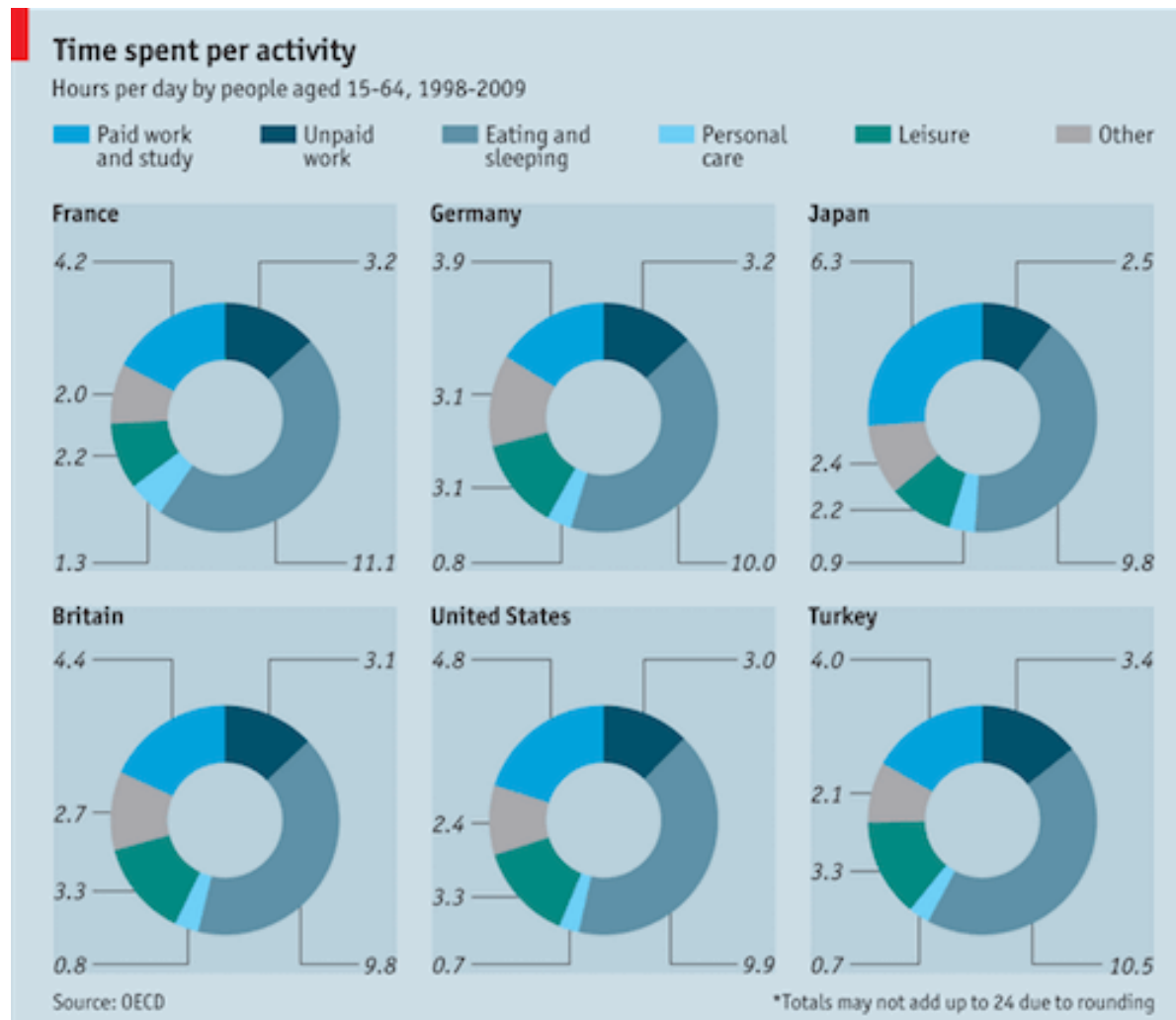
Source: The Economist (via Junk Charts)
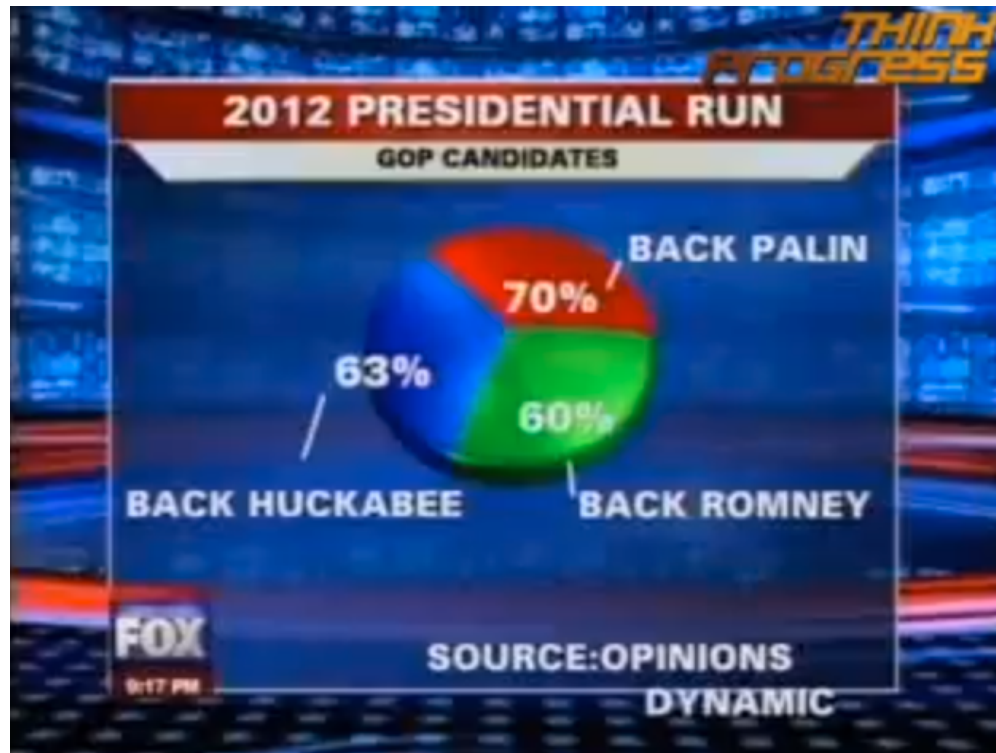
Source: Heap Analytics

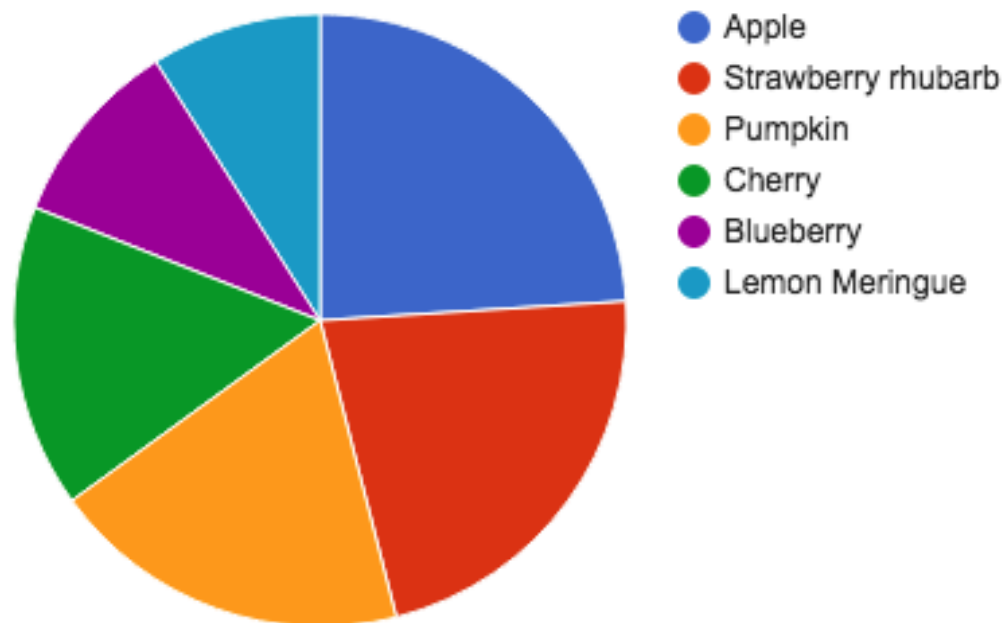# Choosing the right chart …

Let's make a chart (about pie)

Pie type	Percent
Apple	24
Strawberry rhubarb	22
Pumpkin	19
Cherry	16
Blueberry	10
Lemon Meringue	9

Source: NPR

Most popular pie

- Apple — 24%
- Strawberry rhubarb — 22%
- Pumpkin — 19%
- Cherry — 16%
- Blueberry — 10%
- Lemon Meringue — 9%

Most popular pie

- Apple
- Strawberry rhubarb
- Pumpkin
- Cherry
- Blueberry
- Lemon Meringue
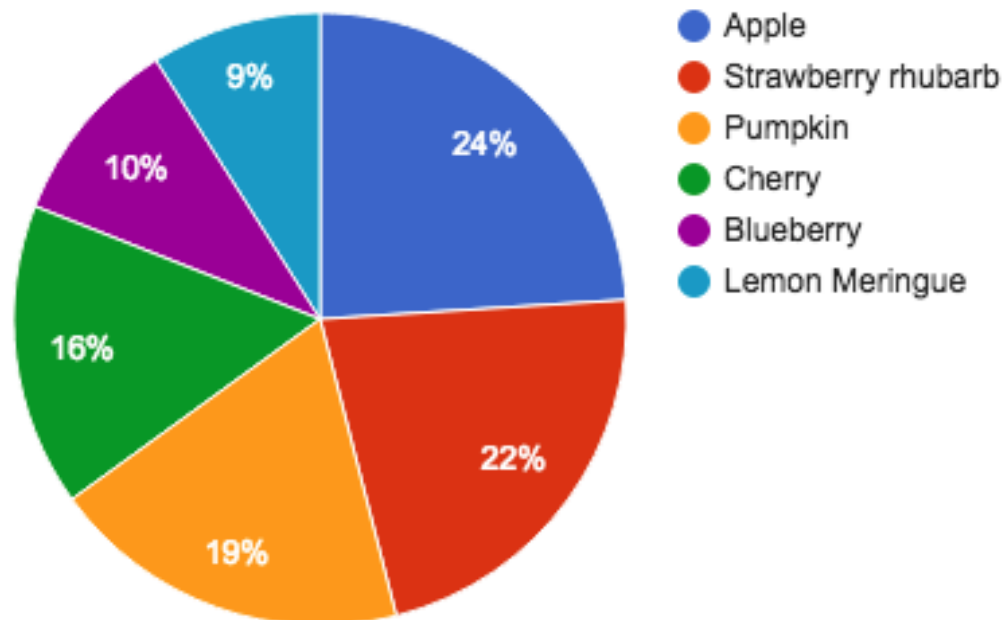
Most popular pie

Most popular pie

Pie type	
Apple	24
Strawberry rhub…	22
Pumpkin	19
Cherry	16
Blueberry	10
Lemon Meringue	9
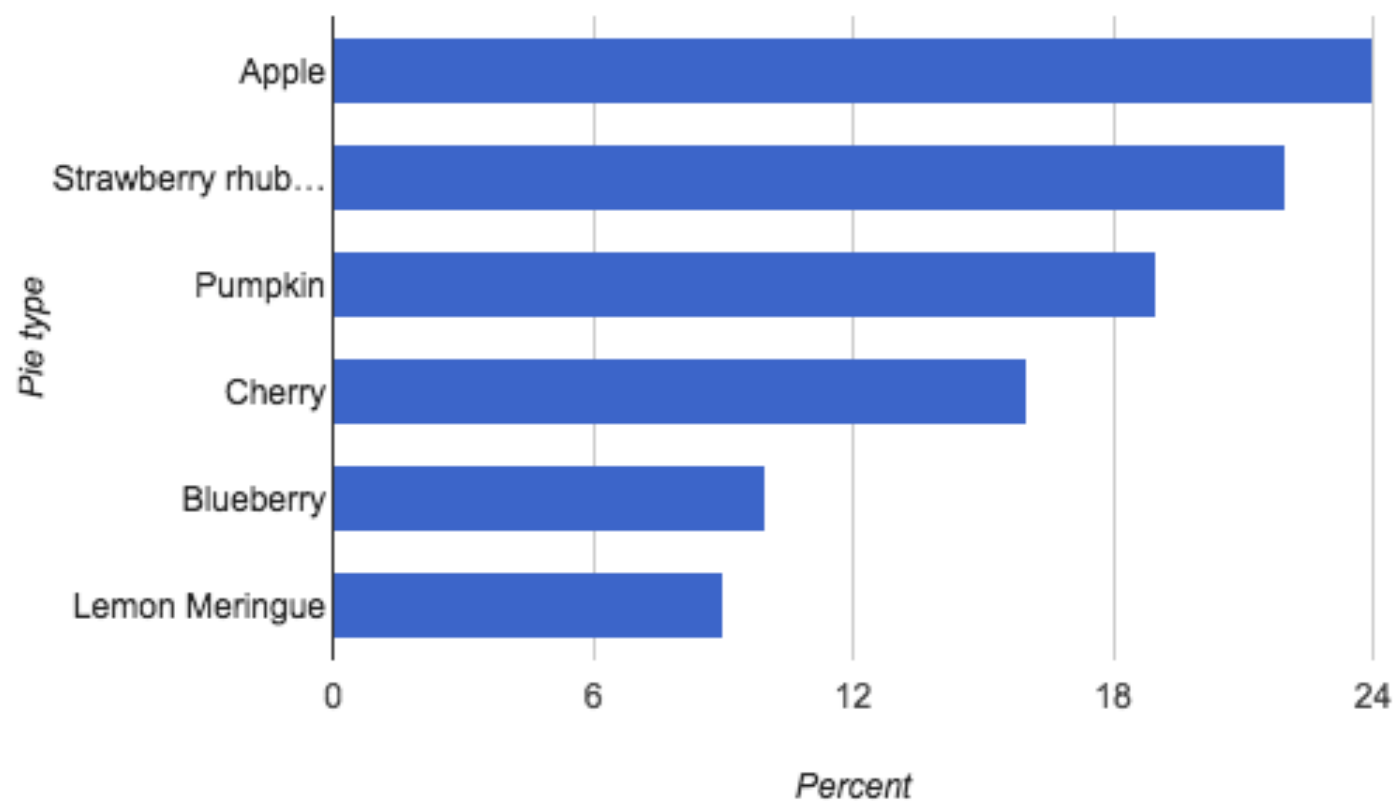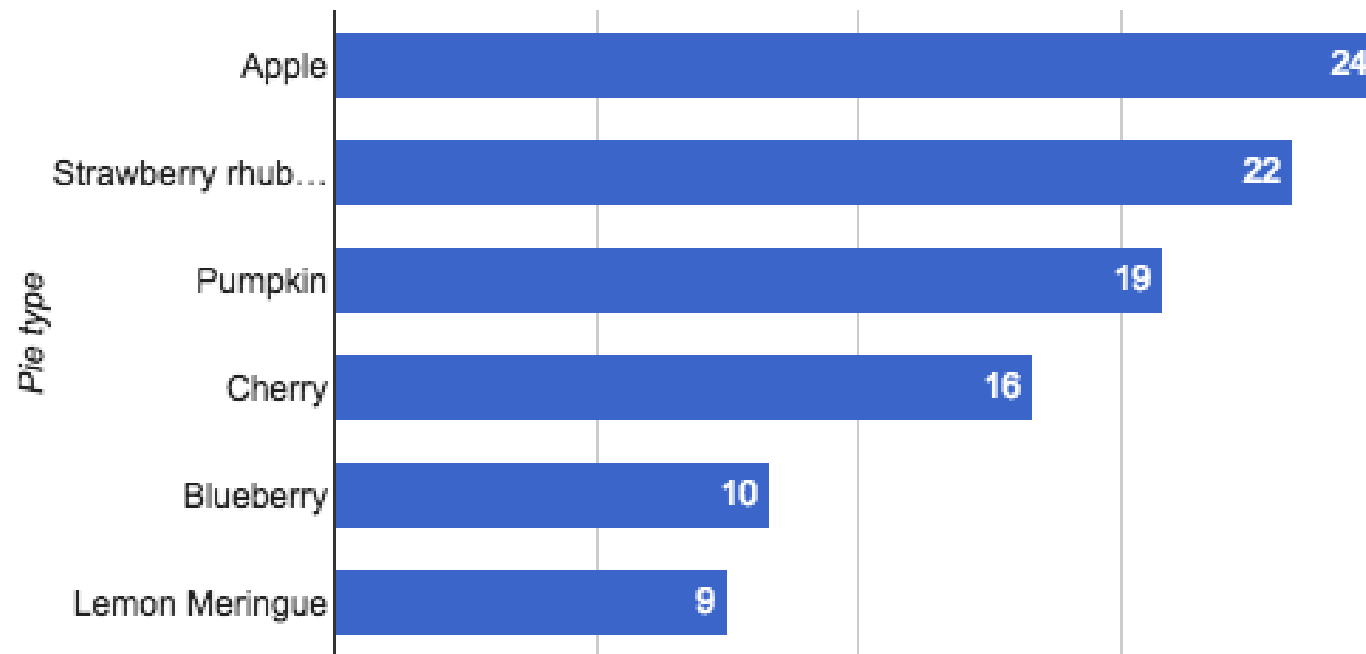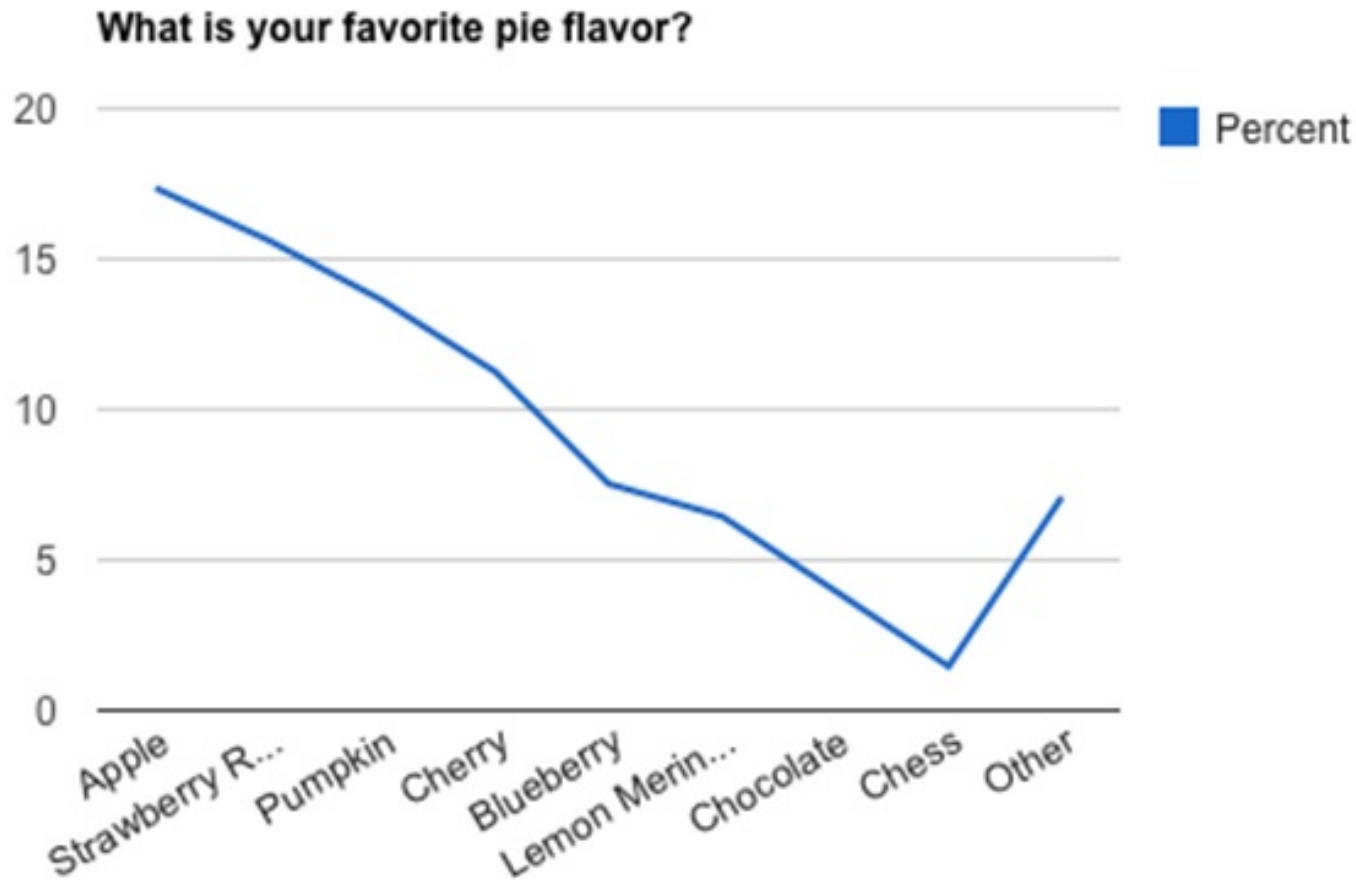
Other chart types to keep in mind ...

A brief overview

Area chart

Long-Term Problem

Share of all unemployed out of work for 27 weeks or longer

Recession

Jan. 2012
42.9%

Source: Bureau of Labor Statistics

The Wall Street Journal
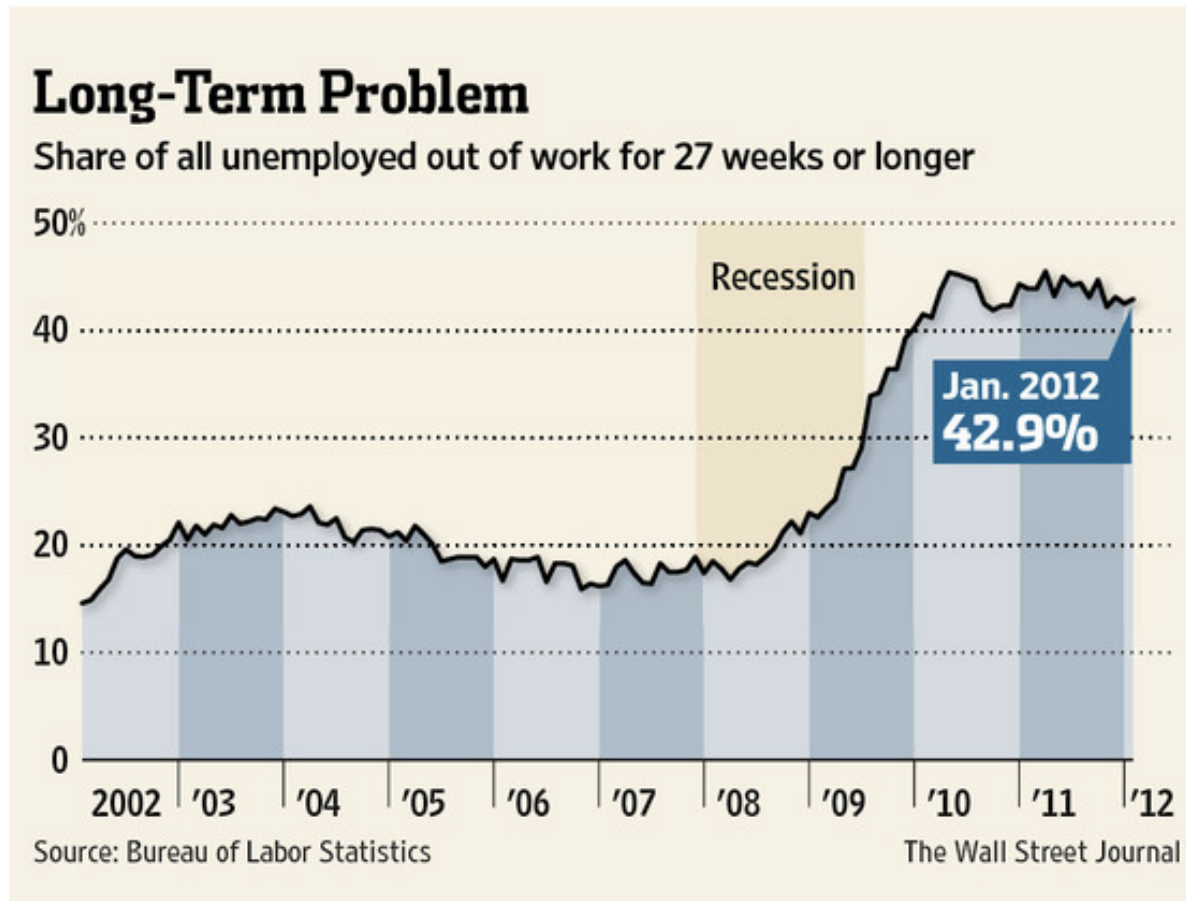
Source: WSJ

How the Speed of Response Defined the Ebola Crisis

By ALBERT SUN and HANNAH FAIRFIELD   NOV. 3, 2014

As the Ebola crisis unfolded in West Africa over the spring and summer, public health experts began calling for speedy international intervention to stop the spread of the disease.

The crucial factor in containing Ebola, these experts say, is to keep infected patients isolated to reduce transmission. If that does not happen, the number of cases shoots up rapidly.

These charts are based on an epidemiological model created by the Centers for Disease Control and Prevention, using figures from Liberia and Sierra Leone, two of the countries hit hardest by the epidemic. In the third, Guinea, case counts have been unreliable.
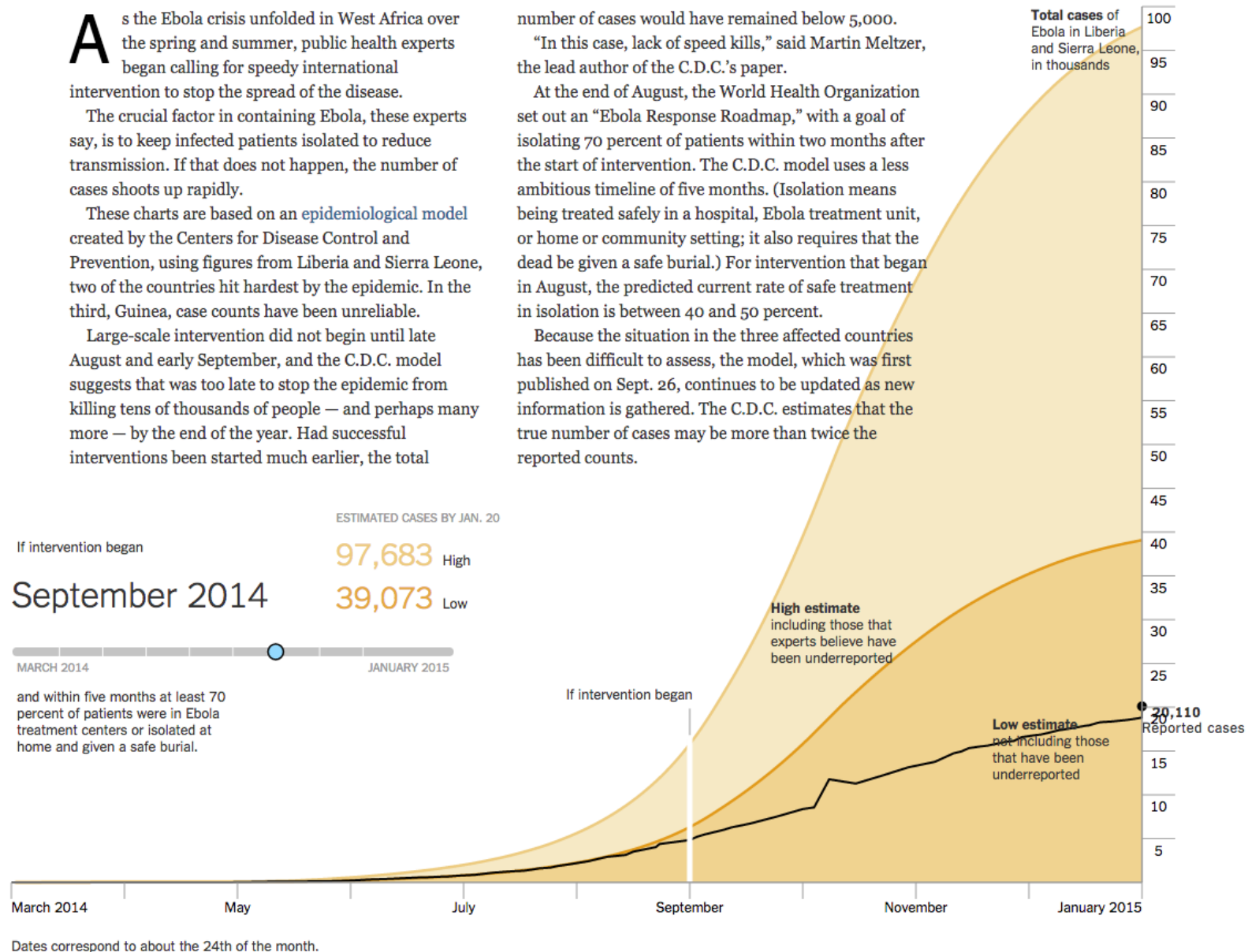
Large-scale intervention did not begin until late August and early September, and the C.D.C. model suggests that was too late to stop the epidemic from killing tens of thousands of people — and perhaps many more — by the end of the year. Had successful interventions been started much earlier, the total number of cases would have remained below 5,000.
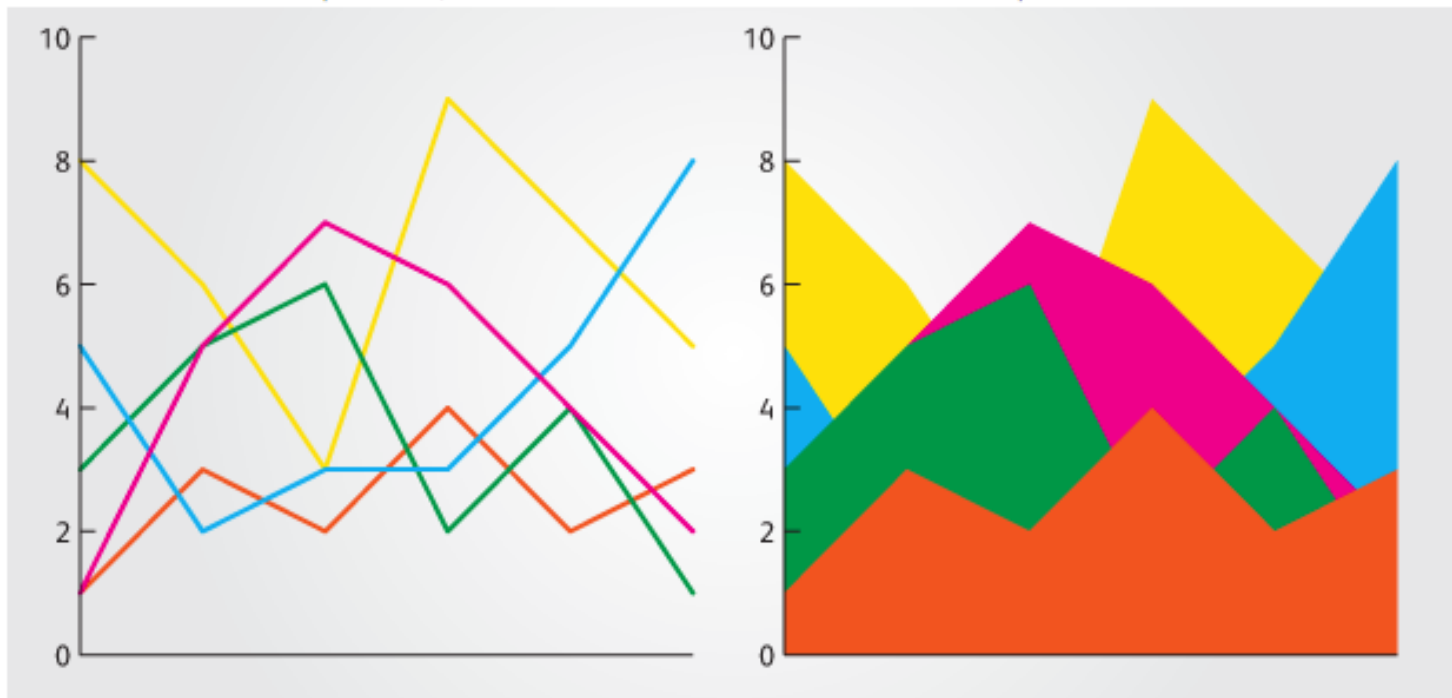
"In this case, lack of speed kills," said Martin Meltzer, the lead author of the C.D.C.'s paper.

At the end of August, the World Health Organization set out an "Ebola Response Roadmap," with a goal of isolating 70 percent of patients within two months after the start of intervention. The C.D.C. model uses a less ambitious timeline of five months. (Isolation means being treated safely in a hospital, Ebola treatment unit, or home or community setting; it also requires that the dead be given a safe burial.) For intervention that began in August, the predicted current rate of safe treatment in isolation is between 40 and 50 percent.

Because the situation in the three affected countries has been difficult to assess, the model, which was first published on Sept. 26, continues to be updated as new information is gathered. The C.D.C. estimates that the true number of cases may be more than twice the reported counts.
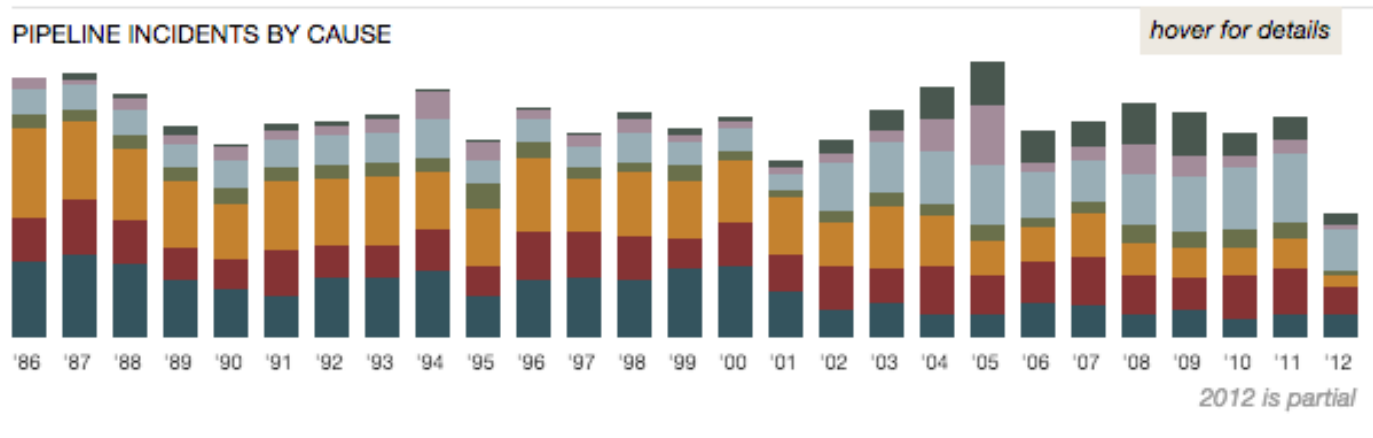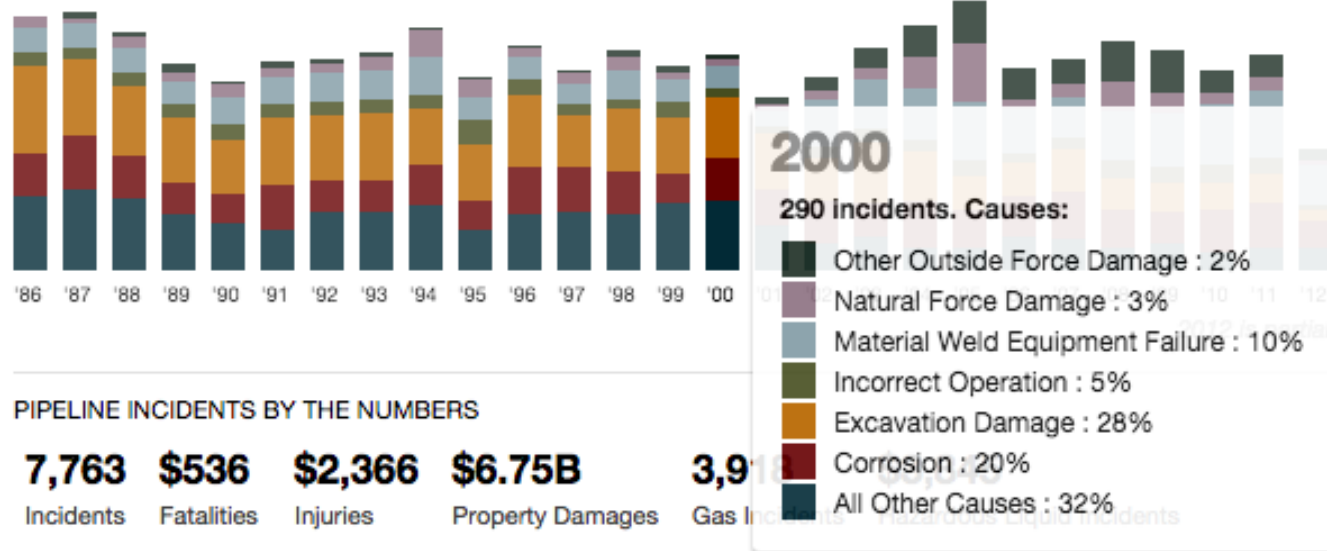
Total cases of Ebola in Liberia and Sierra Leone, in thousands

ESTIMATED CASES BY JAN. 20

If intervention began

September 2014

97,683 High

39,073 Low

MARCH 2014 — JANUARY 2015

and within five months at least 70 percent of patients were in Ebola treatment centers or isolated at home and given a safe burial.

If intervention began

High estimate including those that experts believe have been underreported

Low estimate not including those that have been underreported

20,110 Reported cases

March 2014   May   July   September   November   January 2015

Dates correspond to about the 24th of the month.

Source: New York Times

Source: Visual.ly

Stacked bar chart

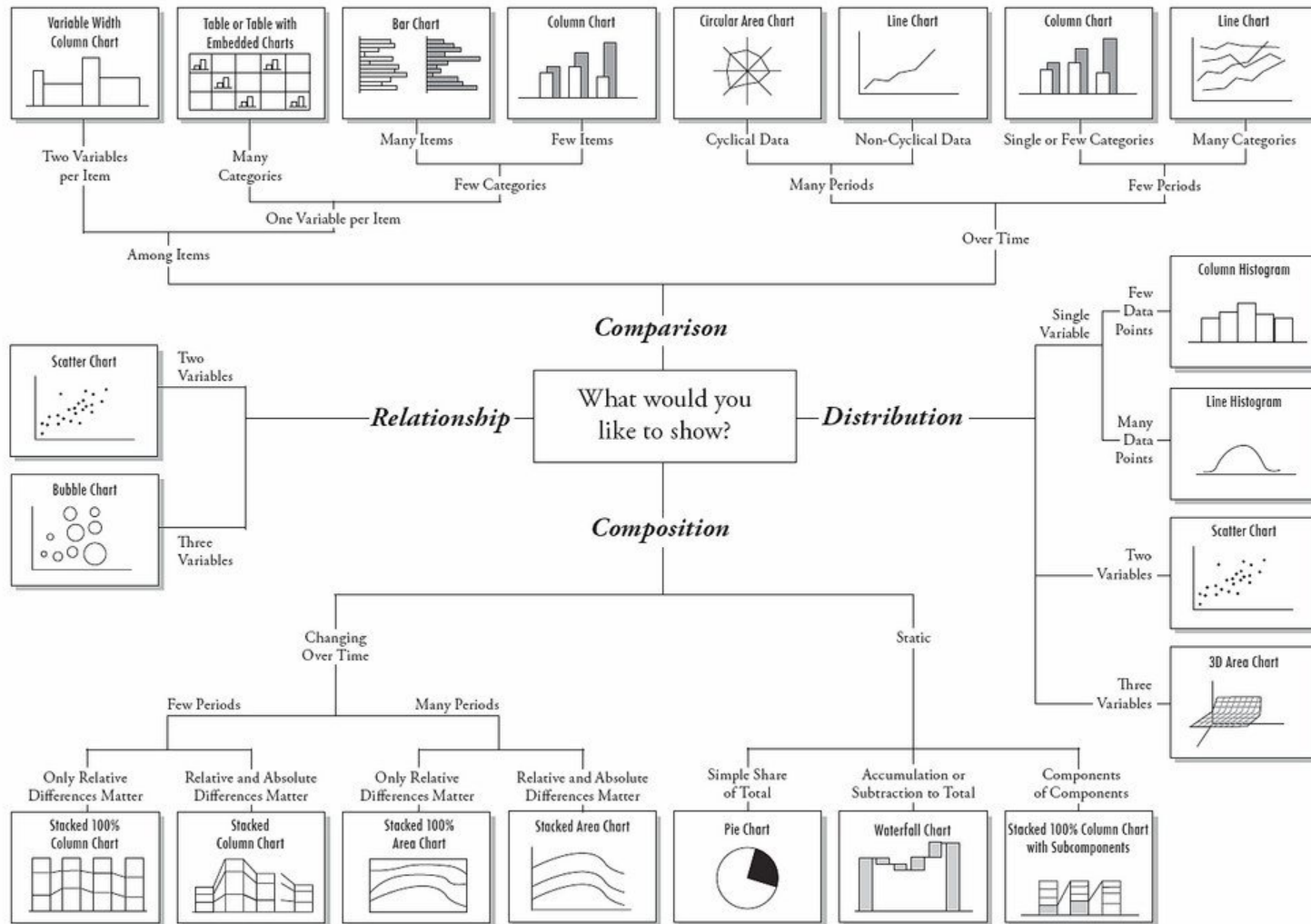Source: ProPublica

PIPELINE INCIDENTS BY CAUSE

hover for details

'86 '87 '88 '89 '90 '91 '92 '93 '94 '95 '96 '97 '98 '99 '00

2000

290 incidents. Causes:

Other Outside Force Damage : 2%
Natural Force Damage : 3%
Material Weld Equipment Failure : 10%
Incorrect Operation : 5%
Excavation Damage : 28%
Corrosion : 20%
All Other Causes : 32%

PIPELINE INCIDENTS BY THE NUMBERS

| **7,763** | **$536** | **$2,366** | **$6.75B** | **3,9** | |
| Incidents | Fatalities | Injuries | Property Damages | Gas Incidents | Hazardous Liquid Incidents |

Source: ProPublica
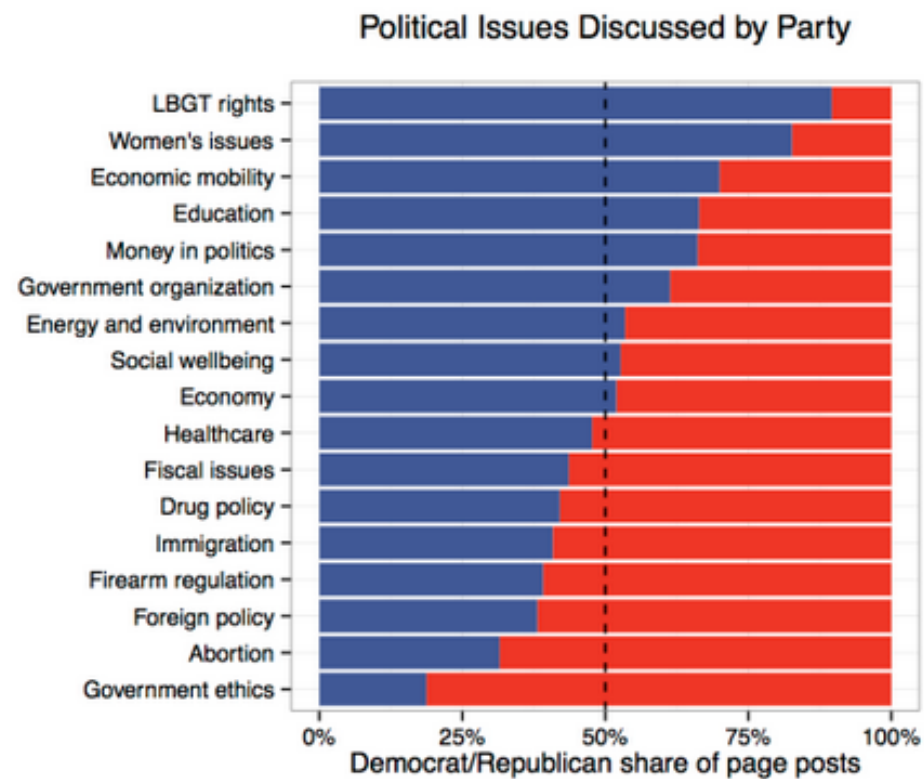
Source: R Bloggers

Chart Suggestions—A Thought-Starter

http://flowingdata.com/2009/01/15/flow-chart-shows-you-what-chart-to-use

Design basics

- Labels and alignment
  - Color
  - Typography

Remove backgrounds
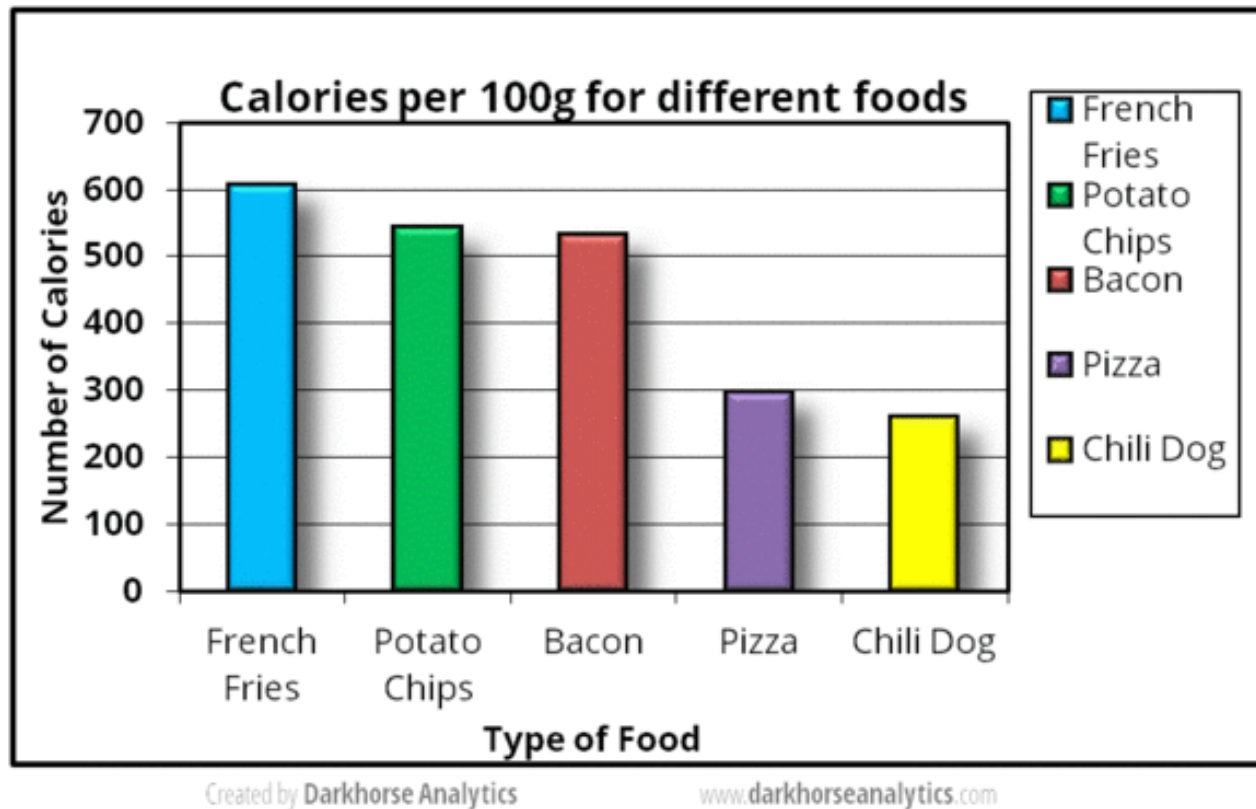
Go here for working gif

Remove backgrounds
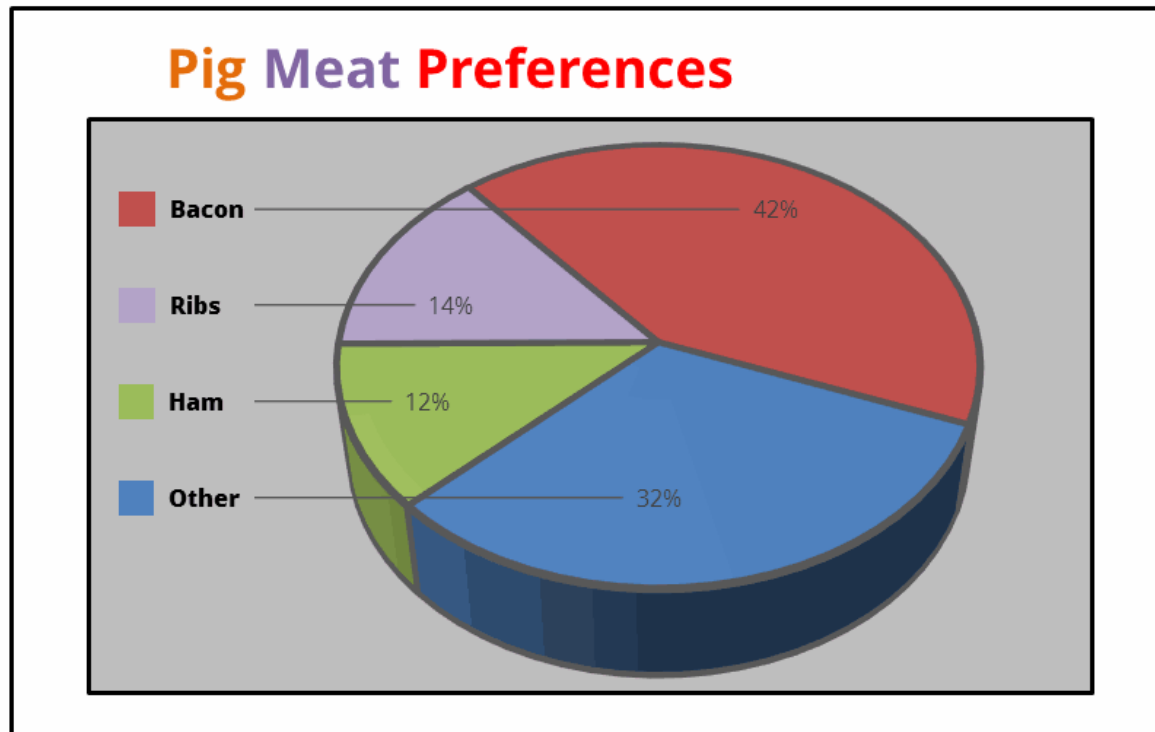
Go here for working gif

Remove the Colors

Role	Name	Year of the...	Debut	Number of Fans	Takedown Rate
Face (The Hero)	The Ultimate Warrior	Tiger	May-2011	97320.00	86.2
Face (The Hero)	Hulk Hogan	Oxen	Jan-2008	988551.00	61.978
Face (The Hero)	Macho Man Randy Savage	Monkey	Feb-2008	157618.00	59.29
Face (The Hero)	Hacksaw Jim Duggan	Pig	Mar-2008	30300.00	53.4332
Face (The Hero)	Superfly Jimmy Snuka	Dragon	Mar-2008	12341.00	52.7
Heel (The Bad Guy)	Rowdy Roddy Piper	Rooster	Jun-1968	71645.00	45.4
Heel (The Bad Guy)	The Million Dollar Man Ted DiBiase	Rat	Apr-1975	449342.00	43.7689
Heel (The Bad Guy)	Mr. Perfect Curt Henning	Rat	May-1980	13773.00	38
Heel (The Bad Guy)	Jake the Snake Roberts	Snake	Jul-1975	5609.00	37.99
Jobber (The Unknown)	Brad Smith	Sheep	Aug-2008	1103.00	36.316
Jobber (The Unknown)	Ted Duncan	Sheep	Aug-2008	200.00	33.61
Jobber (The Unknown)	Joey the Uber Nerd Cherdarchuk	Snake	Aug-2008	5.00	21.0196

Go here for working gif

Chartbuilder

https://quartz.github.io/Chartbuilder/

Slides made using Reveal.js