EMIS/DS 1300: A Practical Introduction to Data Science
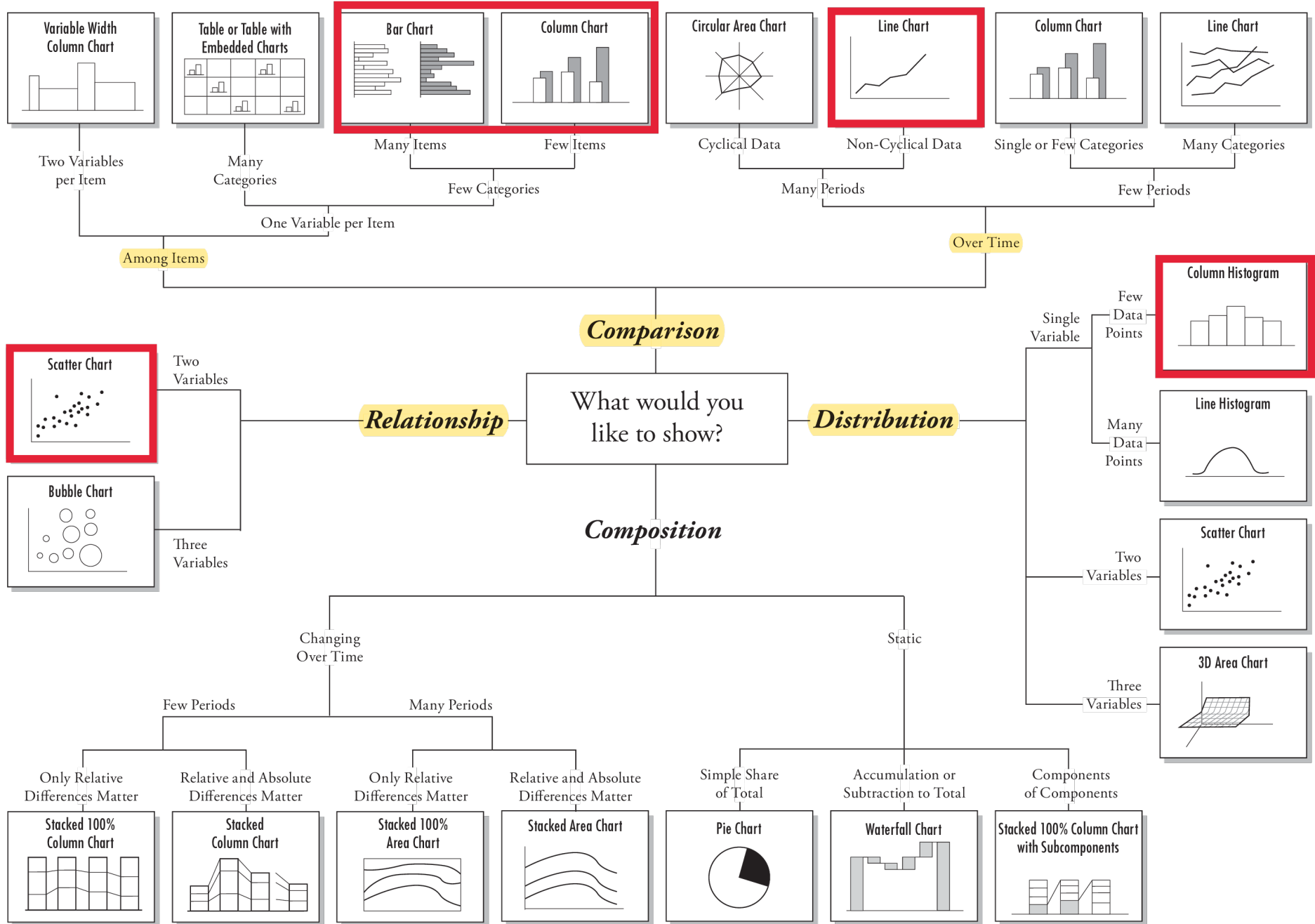
Data Visualization

Slides by Michael Hahsler

Chart Suggestions—A Thought-Starter

Variable Width Column Chart

Two Variables per Item

Table or Table with Embedded Charts

Many Categories

Bar Chart

Many Items

Column Chart

Few Items

Few Categories

One Variable per Item

Among Items

Circular Area Chart

Cyclical Data

Line Chart

Non-Cyclical Data

Many Periods

Column Chart

Single or Few Categories

Line Chart

Many Categories

Few Periods

Over Time

Comparison

Scatter Chart

Two Variables

Relationship

What would you like to show?

Distribution

Single Variable

Column Histogram

Few Data Points

Line Histogram

Many Data Points

Bubble Chart

Three Variables

Composition

Scatter Chart

Two Variables

3D Area Chart

Three Variables

Changing Over Time

Static

Few Periods

Many Periods

Only Relative Differences Matter

Stacked 100% Column Chart

Relative and Absolute Differences Matter

Stacked Column Chart

Only Relative Differences Matter

Stacked 100% Area Chart

Relative and Absolute Differences Matter

Stacked Area Chart

Simple Share of Total

Pie Chart

Accumulation or Subtraction to Total

Waterfall Chart

Components of Components

Stacked 100% Column Chart with Subcomponents

Important Components of Visualization

Components

- **Symbols**: Circles, dots, lines, bars, ...

- **Position**: Axes (labels and units) and the origin (0/0) are important.

- **Length, Size and Area**: Should only be used if it faithfully represents information.

- **Color:** A color should not overpower another. Limit to 3 or 4 colors.

- **Angle:** Human eyes are not good at comparing angles!

Goals

- All components of the plot need to convey information. E.g., do not use color just to make it more colorful.

- Good visualizations show patterns clearly and need little explanation.

Data Visualization

Information visualization is a field of its own.
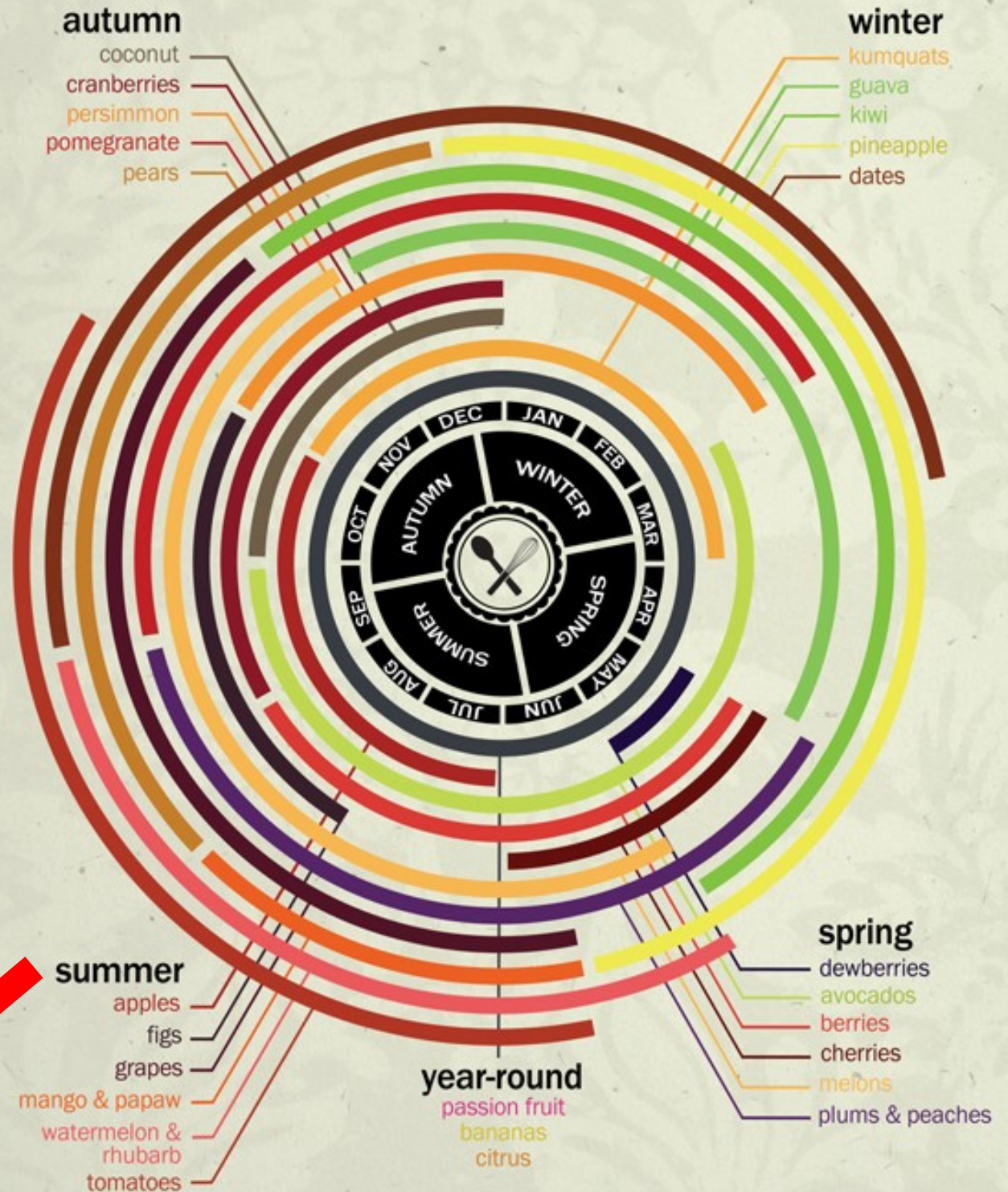
Eat **fruits** when they are in **season!!!**
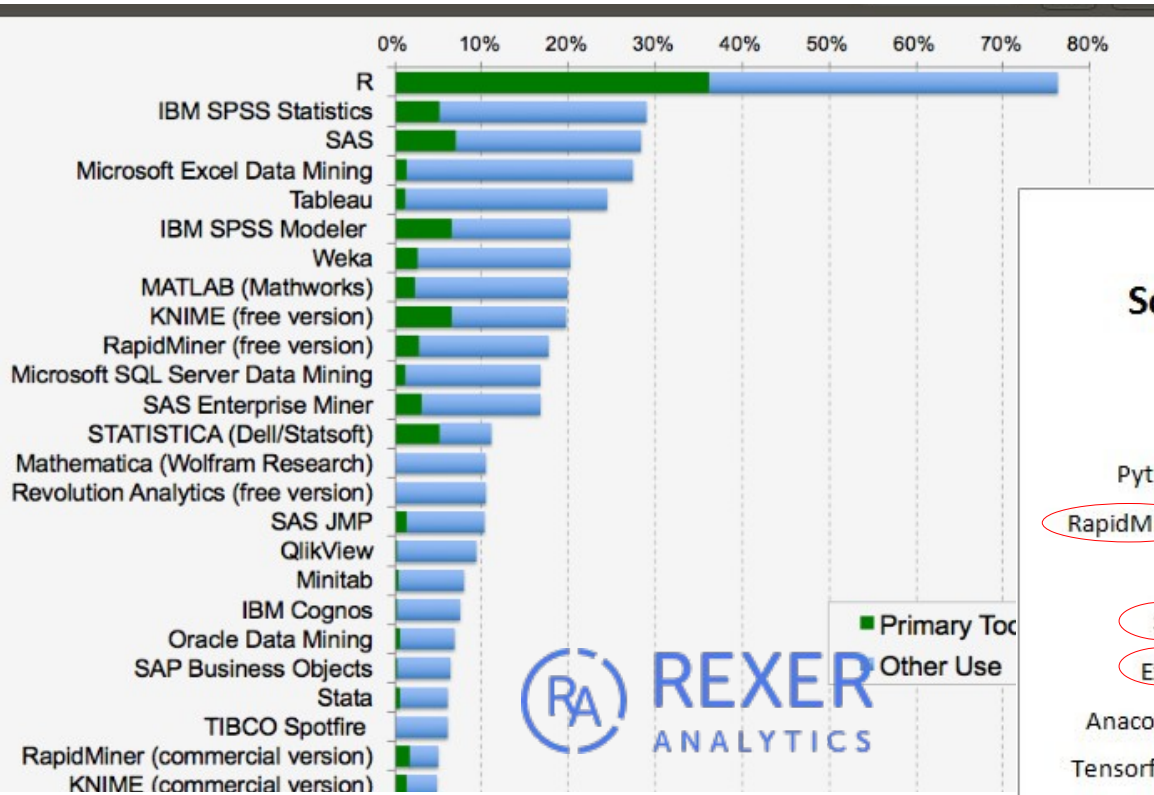
Developmental Aid Flows around the World

■ Top 20 recipients ■ Top 5 donors* • Values in USD million

1 | US 21 787
3 | France 9 884
4 | UK 9 849
2 | Germany 12 291
1 | Iraq 9 115
15 | India 1 298
14 | China 1 439
5 | Japan 7 679

10 | Palestine 1 868
2 | Afghanistan 3 951
13 | Bangladesh 1 502

6 | Pakistan 2 212

19 | Morocco 1 090

18 | Ghana 1 151

4 | Vietnam 2 497

Top 10 US Recipients

Country	Aid Given (USD Million)
Iraq	4 266
Afghanistan	1 459
Sudan	725
Colombia	562
Egypt	541
Congo, Dem. Rep.	486
Pakistan	465
Nigeria	414
Ethiopia	344
Kenya	304

17 | Congo, Dem. Rep. 1 217
16 | Kenya 1 275
20 | Egypt 1083

9 | Cameroon 1 933
3 | Tanzania 2 811
7 | Sudan 2 104

8 | Nigeria 2 042
12 | Uganda 1 728
11 | Mozambique 1 777
5 | Ethiopia 2 422

* Considering only countries donations. European Community donations sum 11 774 USD — standing on the second place in the donors list.

Source: Organisation for Economic Co-operation and Development

Do you notice the slight flaw?

Tools - Popularity

Rexer Analytics 2015

n = 1,220 analytic professionals

http://www.rexeranalytics.com/Data-Miner-Survey-2015-Intro.html
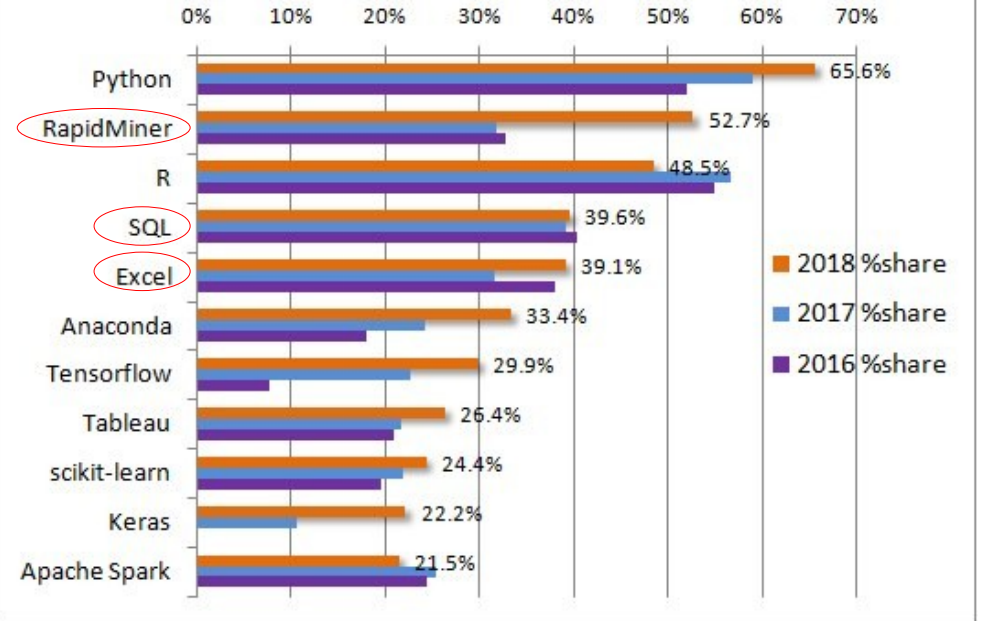
Driving from Kaunas to Moscow

13 hr 18 min
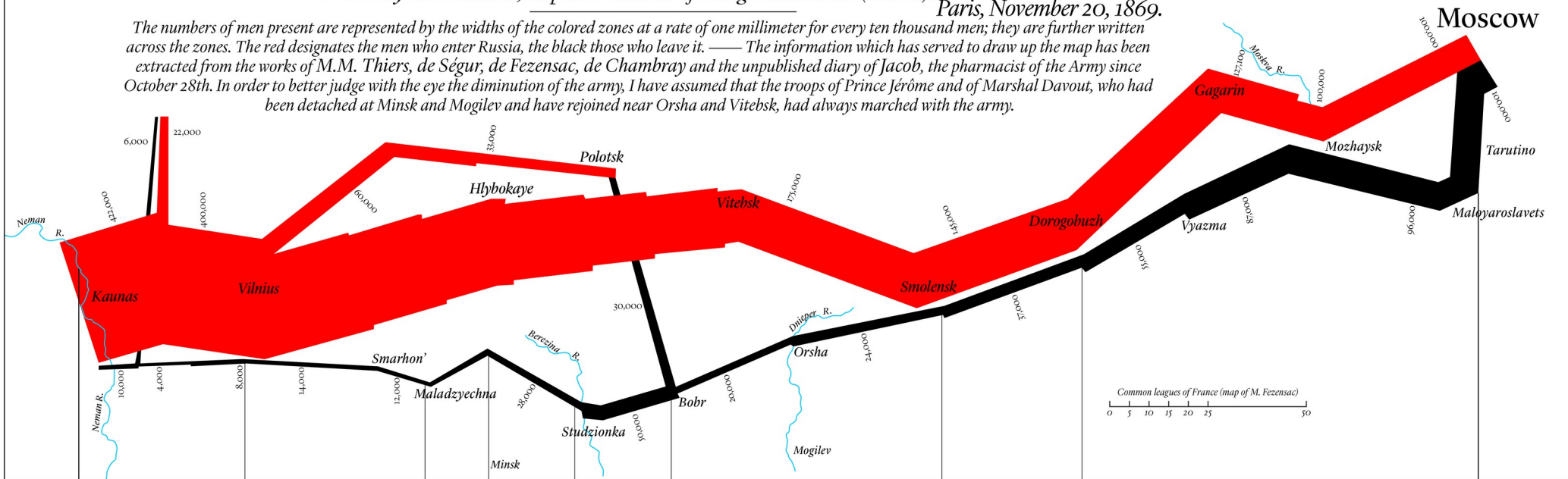979 km
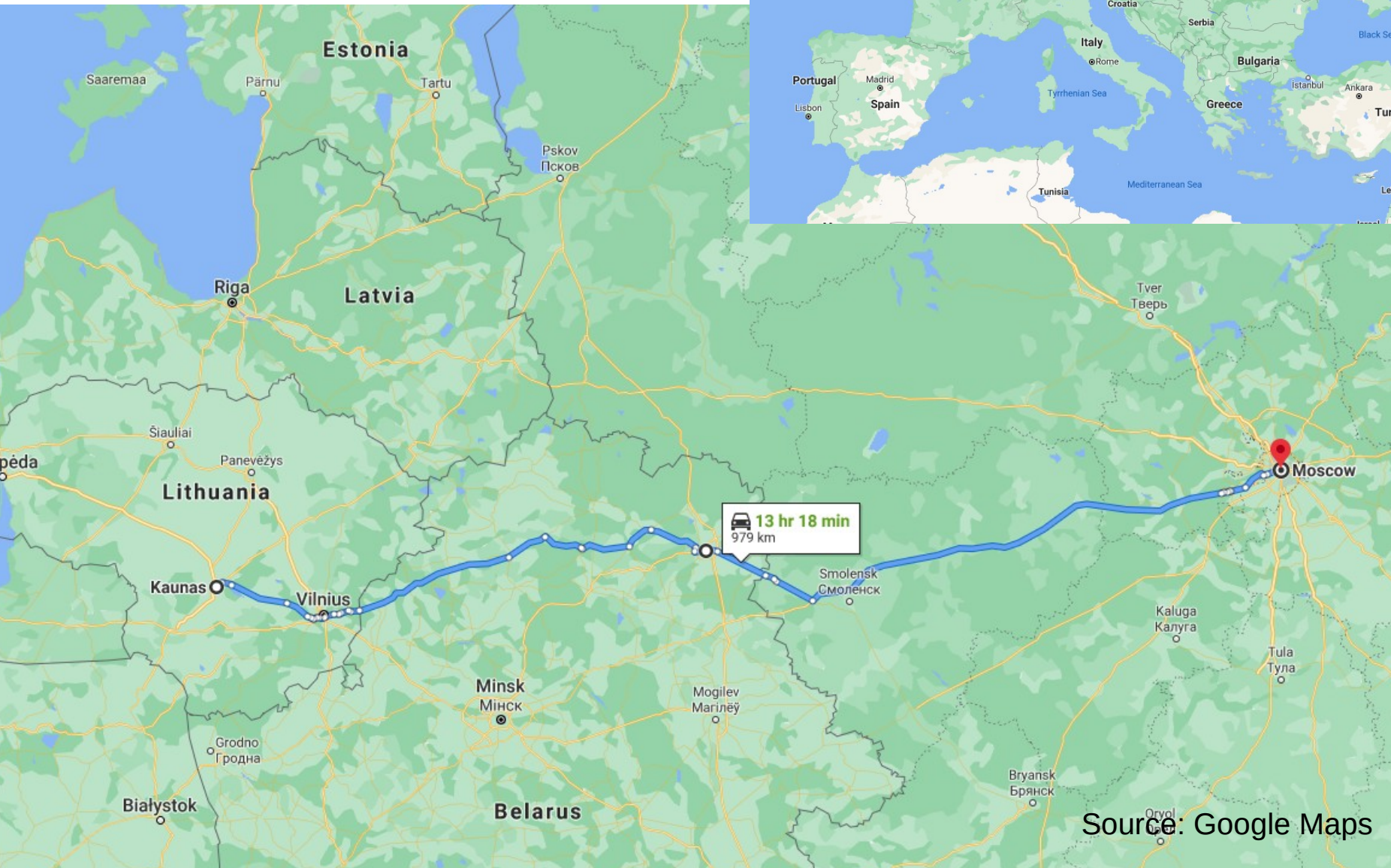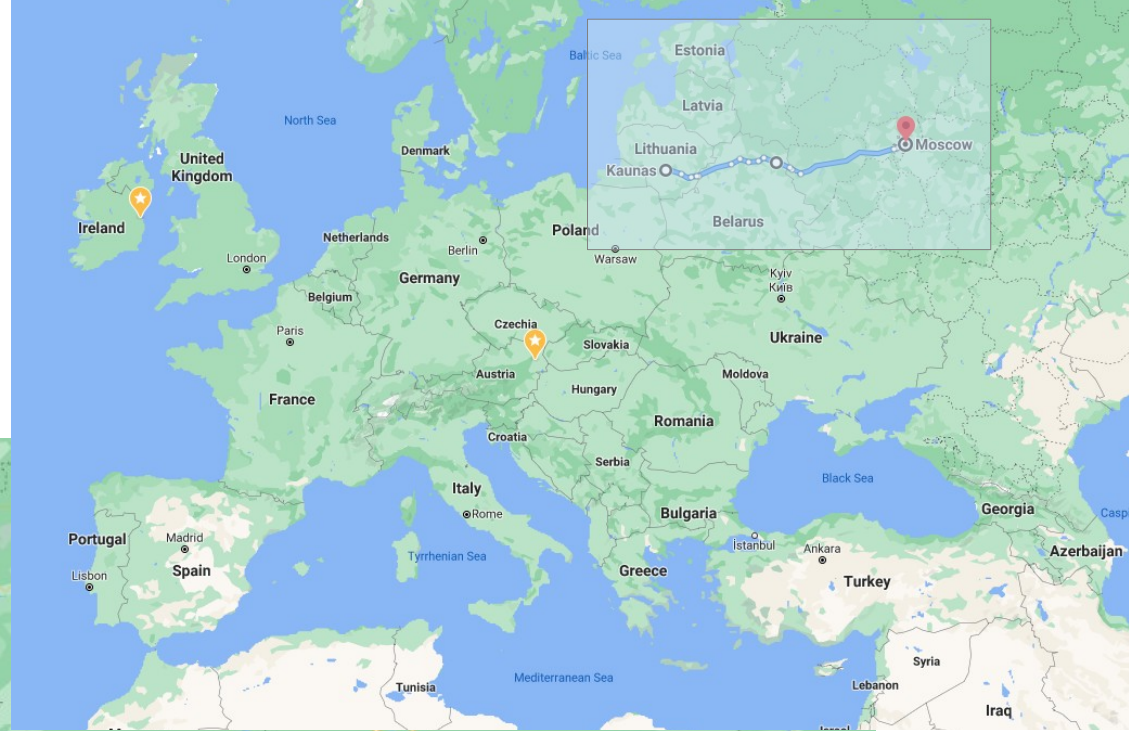
Source: Google Maps

The map of Napoleon's Russian campaign

Figurative Map of the successive losses in men of the *French Army* in the *Russian campaign 1812 ~ 1813*
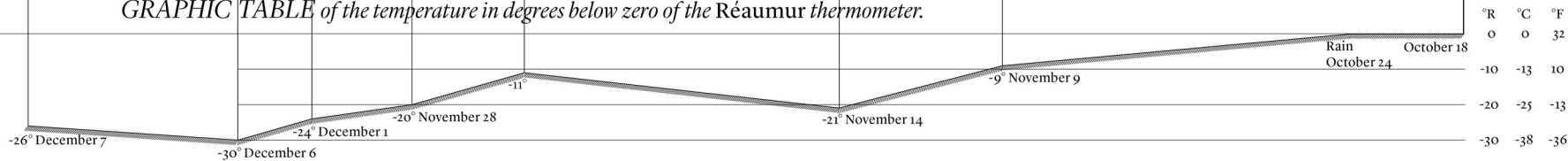Drawn by M. **Minard**, *Inspector General of Bridges and Roads (retired).*
Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. —— The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.

Moscow

Neman R. · Kaunas · Vilnius · Smarhon' · Maladzyechna · Minsk · Studzionka · Berezina R. · Bobr · Mogilev · Orsha · Dnieper R. · Smolensk · Dorogobuzh · Vyazma · Gagarin · Mozhaysk · Tarutino · Maloyaroslavets · Hlybokaye · Polotsk · Vitebsk · Moskva R.

The Cossacks pass the frozen Neman at a gallop.

GRAPHIC TABLE of the temperature in degrees below zero of the Réaumur thermometer.

°R	°C	°F
0	0	32
-10	-13	10
-20	-25	-13
-30	-38	-36

Rain October 18
October 24
-9° November 9
-11°
-20° November 28
-21° November 14
-24° December 1
-26° December 7
-30° December 6

Common leagues of France (map of M. Fezensac)
0   5   10   15   20   25   50

Translated from French (source: https://edspace.american.edu/visualwar/minard/)
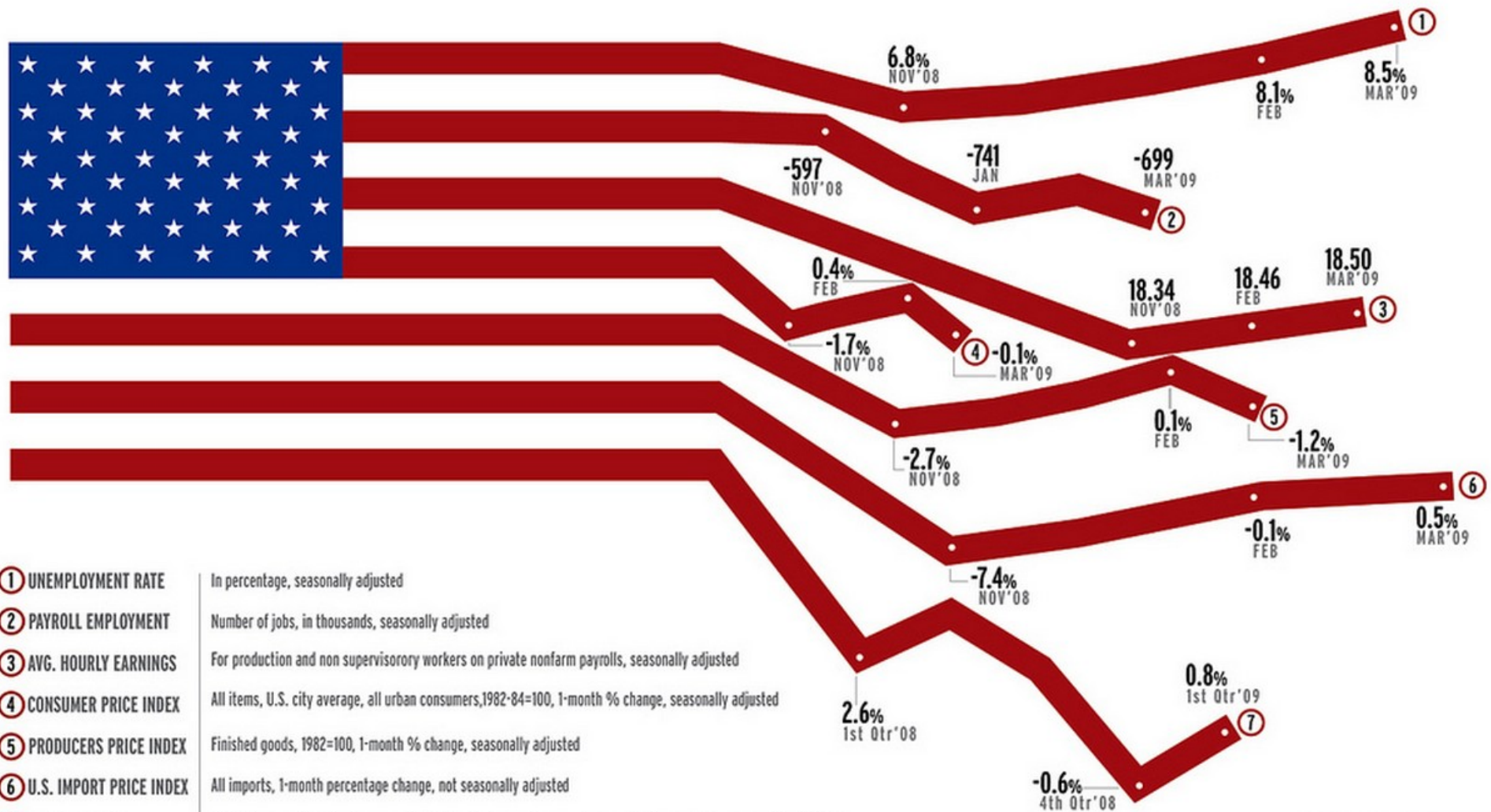
Original created by Charles Joseph Minard, Inspector General of Bridges and Roads (retired).
Paris, November 20, 1869.

Economy at a Glance

The recent failure in the housing and credit markets have resulted in a slowdown in the US economy. 2007 GDP growth was estimated at 2.2% but in 2008 it is projected to be just 0.9%
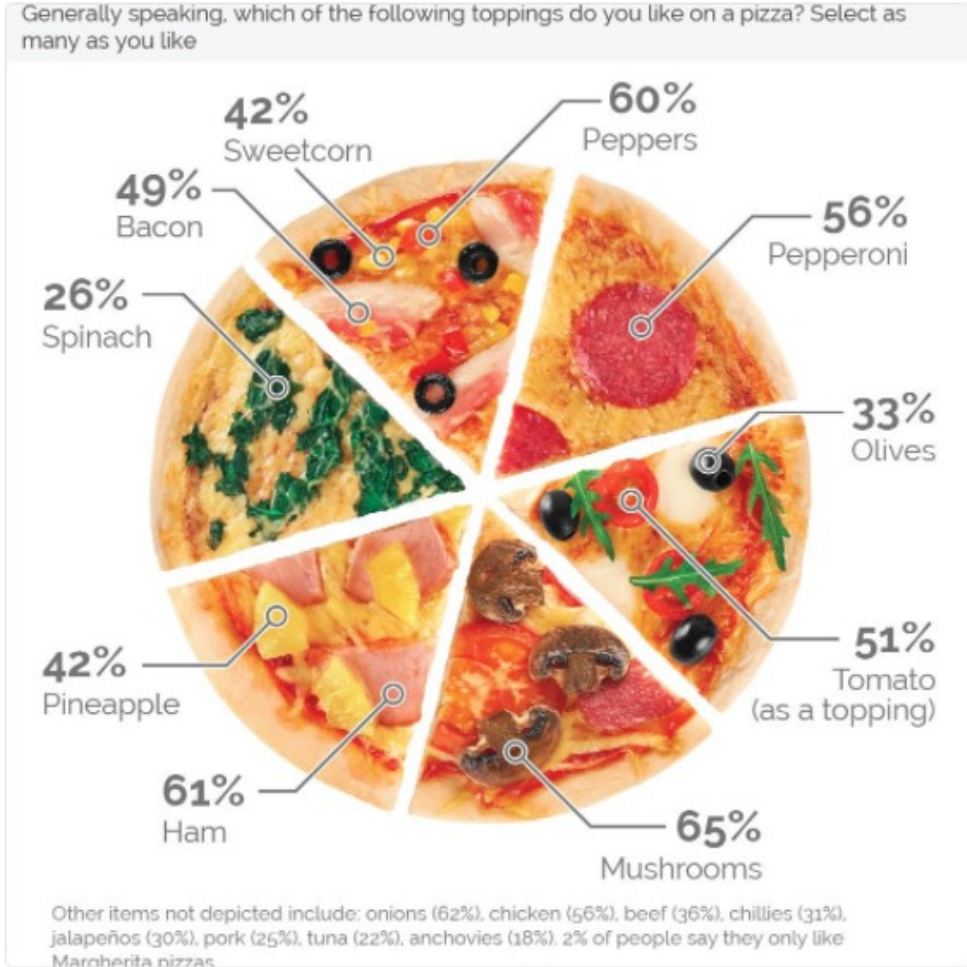
6.8% NOV'08
8.1% FEB
8.5% MAR'09
①

-597 NOV'08
-741 JAN
-699 MAR'09
②

0.4% FEB
18.34 NOV'08
18.46 FEB
18.50 MAR'09
③

-1.7% NOV'08
④ **-0.1%** MAR'09

0.1% FEB
⑤
-1.2% MAR'09

-2.7% NOV'08

-0.1% FEB
⑥
0.5% MAR'09

-7.4% NOV'08

2.6% 1st Qtr'08

0.8% 1st Qtr'09
⑦

-0.6% 4th Qtr'08

① UNEMPLOYMENT RATE	In percentage, seasonally adjusted	
② PAYROLL EMPLOYMENT	Number of jobs, in thousands, seasonally adjusted	
③ AVG. HOURLY EARNINGS	For production and non supervisorory workers on private nonfarm payrolls, seasonally adjusted	
④ CONSUMER PRICE INDEX	All items, U.S. city average, all urban consumers,1982-84=100, 1-month % change, seasonally adjusted	
⑤ PRODUCERS PRICE INDEX	Finished goods, 1982=100, 1-month % change, seasonally adjusted	
⑥ U.S. IMPORT PRICE INDEX	All imports, 1-month percentage change, not seasonally adjusted	
⑦ PRODUCTIVITY	Output/hour, nonfarm business, quarterly data,% change from previous quater at annual rate, seasonally adjusted	

Source: Bureau of labour statistic Graphics: **RAJ**

Fox News – 2012 presidential race
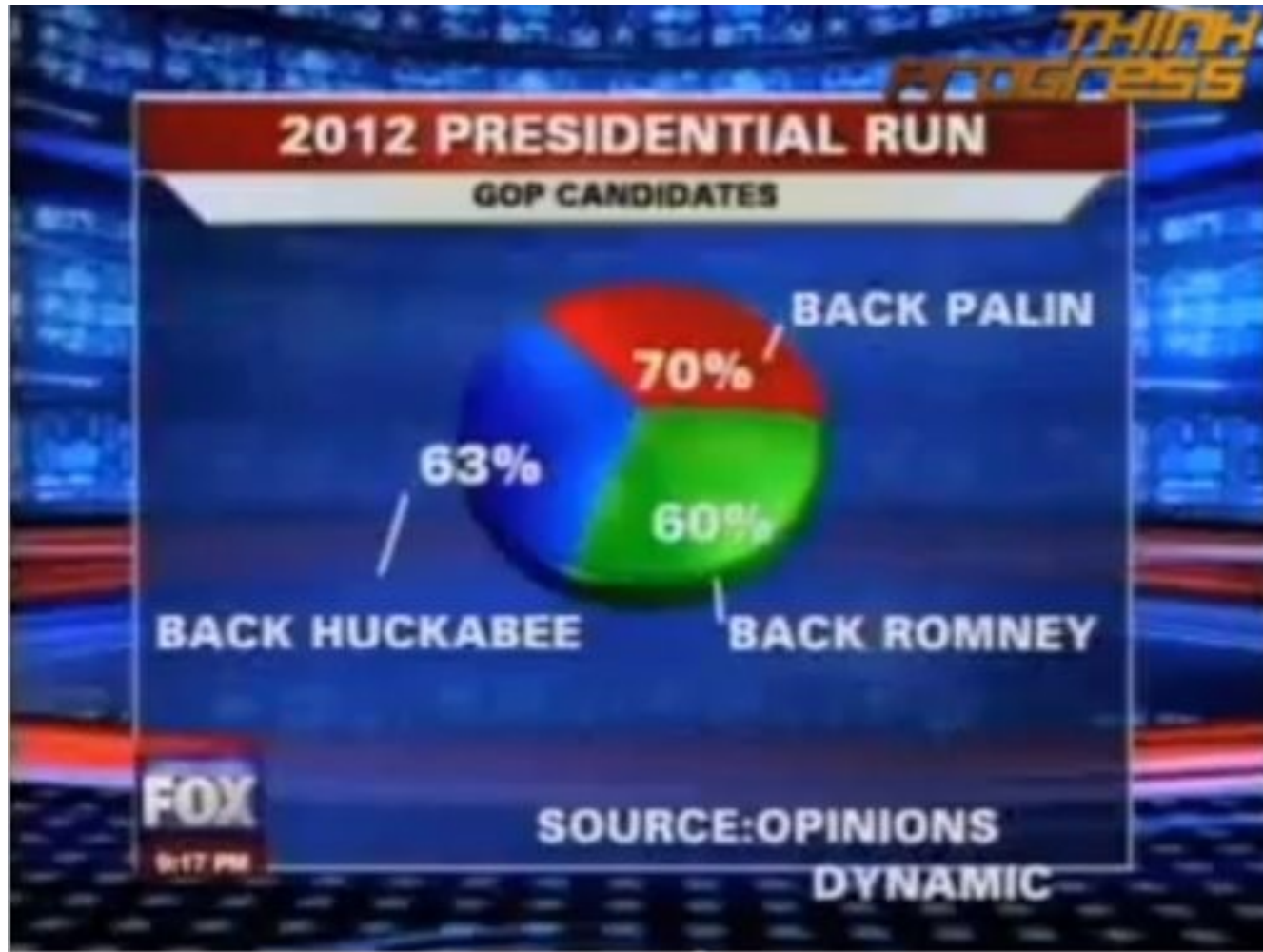
Forget pepperoni - mushroom is Britain's
most liked pizza topping (65%), followed by
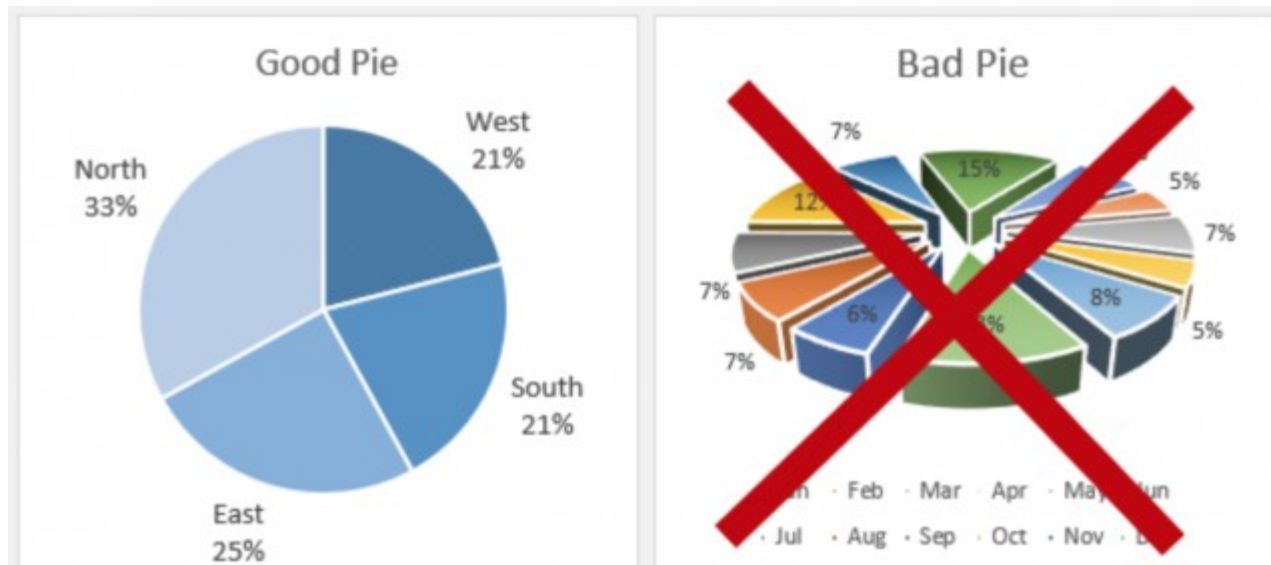onion (62%) and then ham (61%)
yougov.co.uk/news/2017/03/0 ...

Generally speaking, which of the following toppings do you like on a pizza? Select as
many as you like

42% Sweetcorn
60% Peppers
49% Bacon
56% Pepperoni
26% Spinach
33% Olives
42% Pineapple
51% Tomato (as a topping)
61% Ham
65% Mushrooms

Other items not depicted include: onions (62%), chicken (56%), beef (36%), chillies (31%),
jalapeños (30%), pork (25%), tuna (22%), anchovies (18%). 2% of people say they only like
Margherita pizzas

4:00 AM - 6 Mar 2017

364 Retweets 549 Likes

179   364   549

Is this pie a pie chart?

You could use a bar chart instead
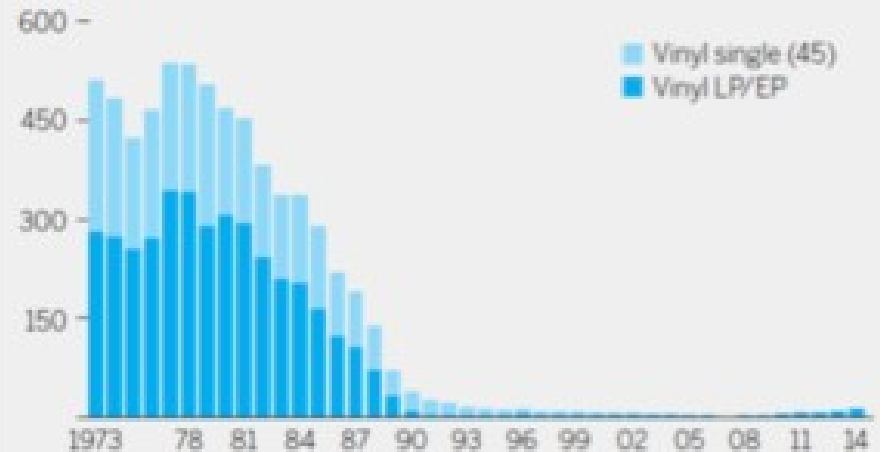
Axes

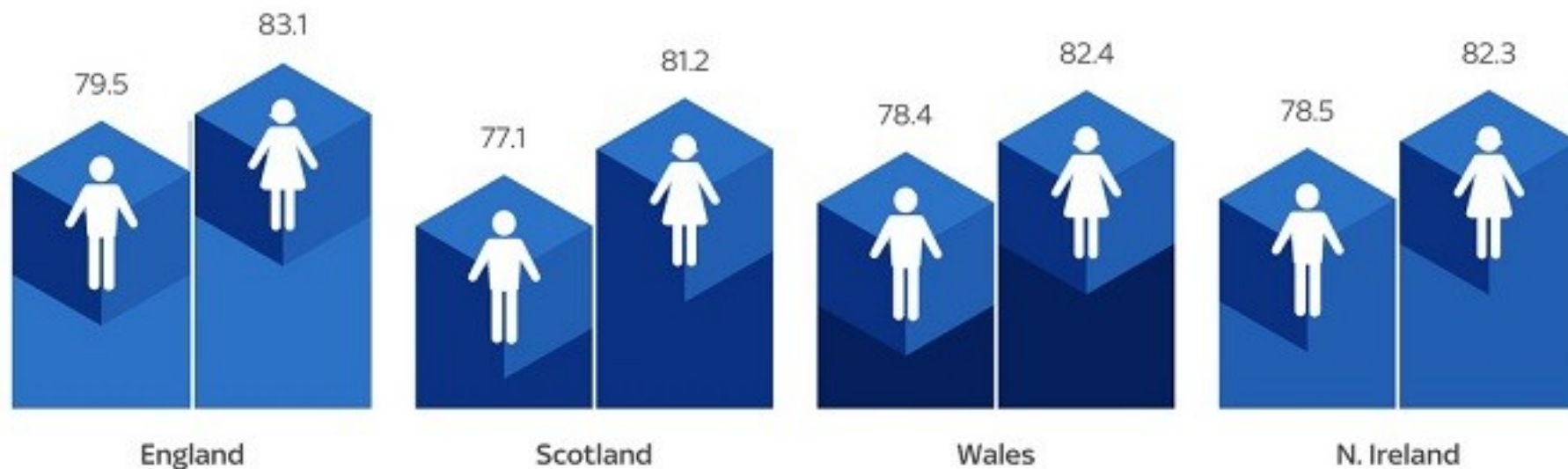**VINYL SALES SINCE 1993**

MILLIONS OF UNITS
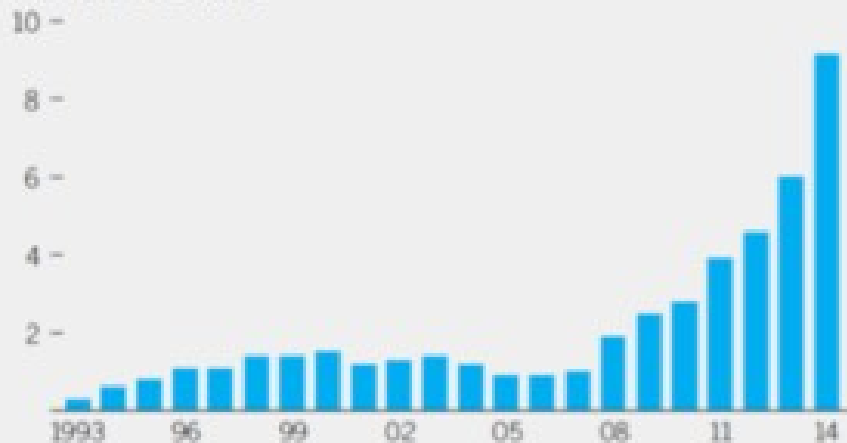
SOURCE: RIAA

VINYL SALES SINCE 1973

MILLIONS OF UNITS

- Vinyl single (45)
- Vinyl LP/EP

Sky News – Average UK life expectancy

81.2 Average UK life expectancy

	England		Scotland		Wales		N. Ireland	
	79.5	83.1	77.1	81.2	78.4	82.4	78.5	82.3

UK Life Expectancy (fixed)