Discussion #5

Name:

Data Visualization

United States $7,290

Dollar figures reflect all public and private spending on care, from doctor visits to hospital infrastructure. Data are from 2007 or the most recent year available.

GRAPHIC: OLIVER UBERTI, NG STAFF. SOURCE: "OECD HEALTH DATA 2009," ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Health care spending per person, in U.S. dollars

Average number of doctor visits a year

0   4   8   12+

— Nation with universal health coverage provided by public and private insurers

— Nation without universal health coverage

Average life expectancy at birth

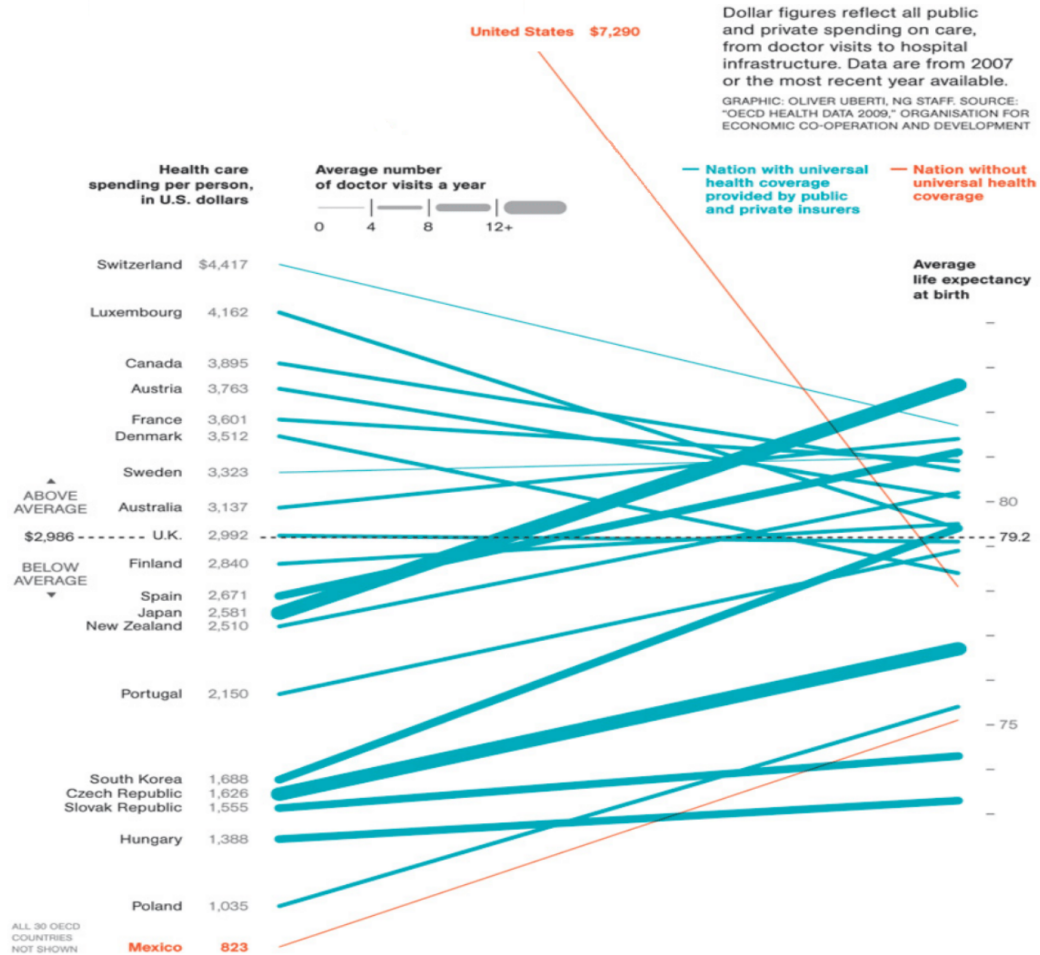| Country	Spending
Switzerland	$4,417
Luxembourg	4,162
Canada	3,895
Austria	3,763
France	3,601
Denmark	3,512
Sweden	3,323
Australia	3,137
U.K.	2,992
Finland	2,840
Spain	2,671
Japan	2,581
New Zealand	2,510
Portugal	2,150
South Korea	1,688
Czech Republic	1,626
Slovak Republic	1,555
Hungary	1,388
Poland	1,035
Mexico	823

▲ ABOVE AVERAGE
$2,986
BELOW AVERAGE ▼

— 80
79.2

— 75

ALL 30 OECD COUNTRIES NOT SHOWN

1. The first part of the discussion will be centered on the above visualization.

   (a) Five variables are being represented visually in this graphic. What are they and what are their types?

(b) How are the variables represented in the graphic, e.g., the variable XXX is mapped to the x-axis, the variable WWW is mapped to the y-axis, the variable ZZZ is conveyed through color, etc.?

(c) How can we figure out how to interpret the visual qualities of the plot, e.g., how do we know what a color represents?

(d) What purpose does the comment at the top right of the plot serve?

(e) Make 3 observations about the figure. Describe the feature that you are basing your observation on.

For example, South Korea's expenditure on health care is comparable to Eastern European countries (and among the lowest of all countries plotted), but the life expectancy is much higher than the Eastern European countries. In the plot we see that the left endpoint of South Korea's line segment is near the Eastern European countries, but the slope of the line segment is much steeper.

(f) Consider the steep negative slope and narrowness of the line segment that represents the data for the United States. What systemic, social, or societal issues might explain this?

2. Name and sketch some appropriate printed (on paper) 2D visualizations if your goal is to explore:
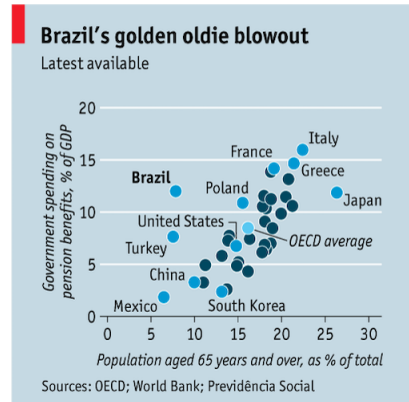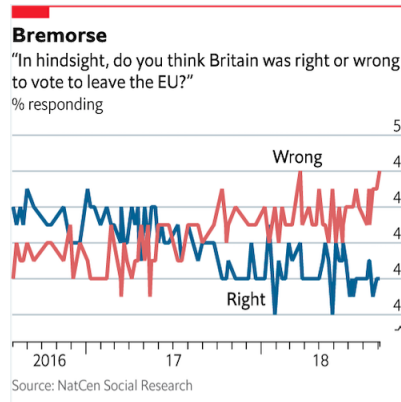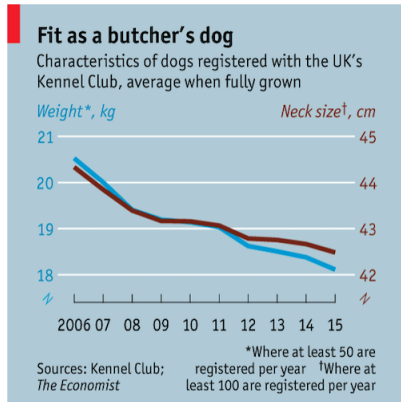
    (a) The distribution of population for various cities.

    (b) The distribution of income.

    (c) The relationship between income and life expectancy.

    (d) The relationship between income and city.

    (e) The relationship between income, life expectancy, smoking status(non-smoker, social smoker, smoker), and city

3. Creating visualizations that represent data accurately and that support the narrative we wish to create is no easy task. Even the journalists and editors at *The Economist*, a newspaper known for it's compelling, data-driven articles, have been known to make blunders. Three of their ill-thought-out plots are presented below. Consider what aspects of the visualizations are misleading, and think of ways in which you can remedy them.

Fit as a butcher's dog
Characteristics of dogs registered with the UK's Kennel Club, average when fully grown

Weight*, kg   *Neck size†, cm*

Sources: Kennel Club; The Economist

*Where at least 50 are registered per year   †Where at least 100 are registered per year

Bremorse
"In hindsight, do you think Britain was right or wrong to vote to leave the EU?"
% responding

Wrong

Right

Source: NatCen Social Research

Brazil's golden oldie blowout
Latest available

Government spending on pension benefits, % of GDP

Italy
France
Greece
Brazil
Poland
Japan
United States
OECD average
Turkey
China
Mexico
South Korea

Population aged 65 years and over, as % of total

Sources: OECD; World Bank; Previdência Social

Hint: The datapoints in the rightmost plot are shaded based on whether or not they are labeled.