

# Data Visualization

INFO6540

Week 3 - Jan 23, 2018

# Assignment 1

Since the second part of the GitHub lab will take place after Assignment 1 is due, please submit your files for it via Brightspace.

# Week 2 lab - Part 2

- February 6, after the guest lecture by Dr. Finlay Maguire
- GitHub tutorial using a GUI
- We will use GitKraken
- Before the lab, go to <https://www.gitkraken.com/> and install it
- Links to tutorials for GitKraken have been added to Brightspace (Week 2)

# Week 3 lab - preview

Download and install Protege Desktop

<https://protege.stanford.edu/>

Do it before class next week!

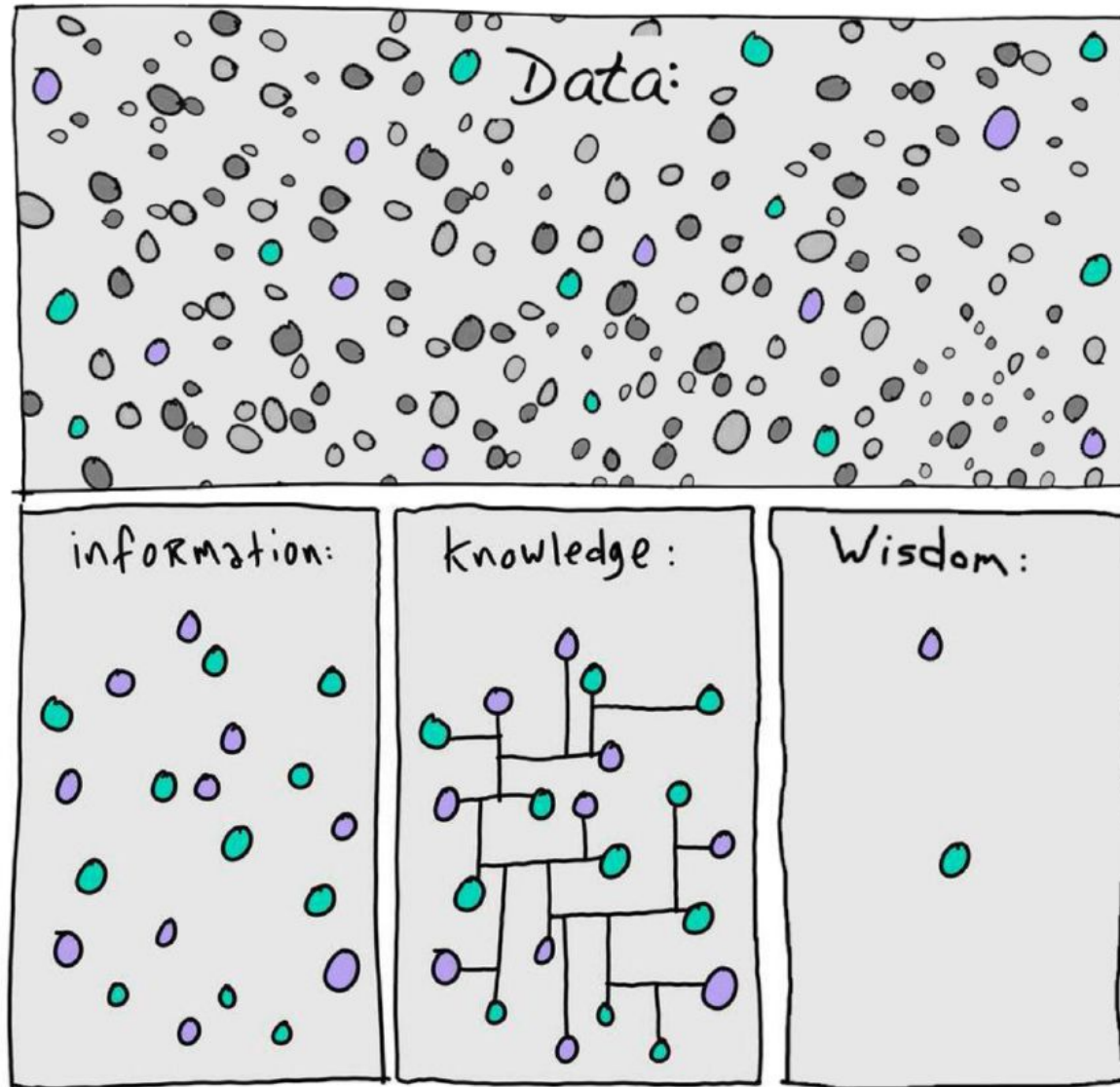
# Week 2

- Data sharing
- Version control

# Visualization

Using graphics to climb the Data – Information – Knowledge –  
Wisdom hierarchy.

# Visualization



**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

# Two main reasons to visualize data

- Exploration
- Communication



# How do you visualize data?

- Decide intention
  - What do you wish to communicate or learn?
- Choose appropriate visualization
  - (e.g., scatter plot, bar chart, histogram, etc.)
- Choose the right tool

# Choosing a Visualization

- What type of data do you have?
- What do you want to communicate or learn?
- Who is your audience?

# Advantages of visualizing data

- Point to important information
- Summarize information
- Help understanding and interpretation
- Draw attentions to supporting facts

# Revenue Trends

	Q1	Q2	Q3	Q4	Total
Americas	-18%	7%	25%	2%	2%
Australia	47%	-7%	26%	15%	17%
China	15%	-5%	1%	7%	19%
Europe	57%	10%	-3%	7%	13%
India	57%	6%	-3%	8%	13%

**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

# Revenue Trends

	Q1	Q2	Q3	Q4	Total
Americas	-18%	7%	25%	2%	2%
Australia	47%	-7%	26%	15%	17%
China	15%	-5%	1%	7%	19%
Europe	57%	10%	-3%	7%	13%
India	57%	6%	-3%	8%	13%

**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management



- International Stocks
- Large-Cap U.S. Stock
- Bonds
- Real Estate
- Mid-Cap U.S. Stock
- Small-Cap U.S. Stock
- Commodities

## Investment Portfolio Breakdown

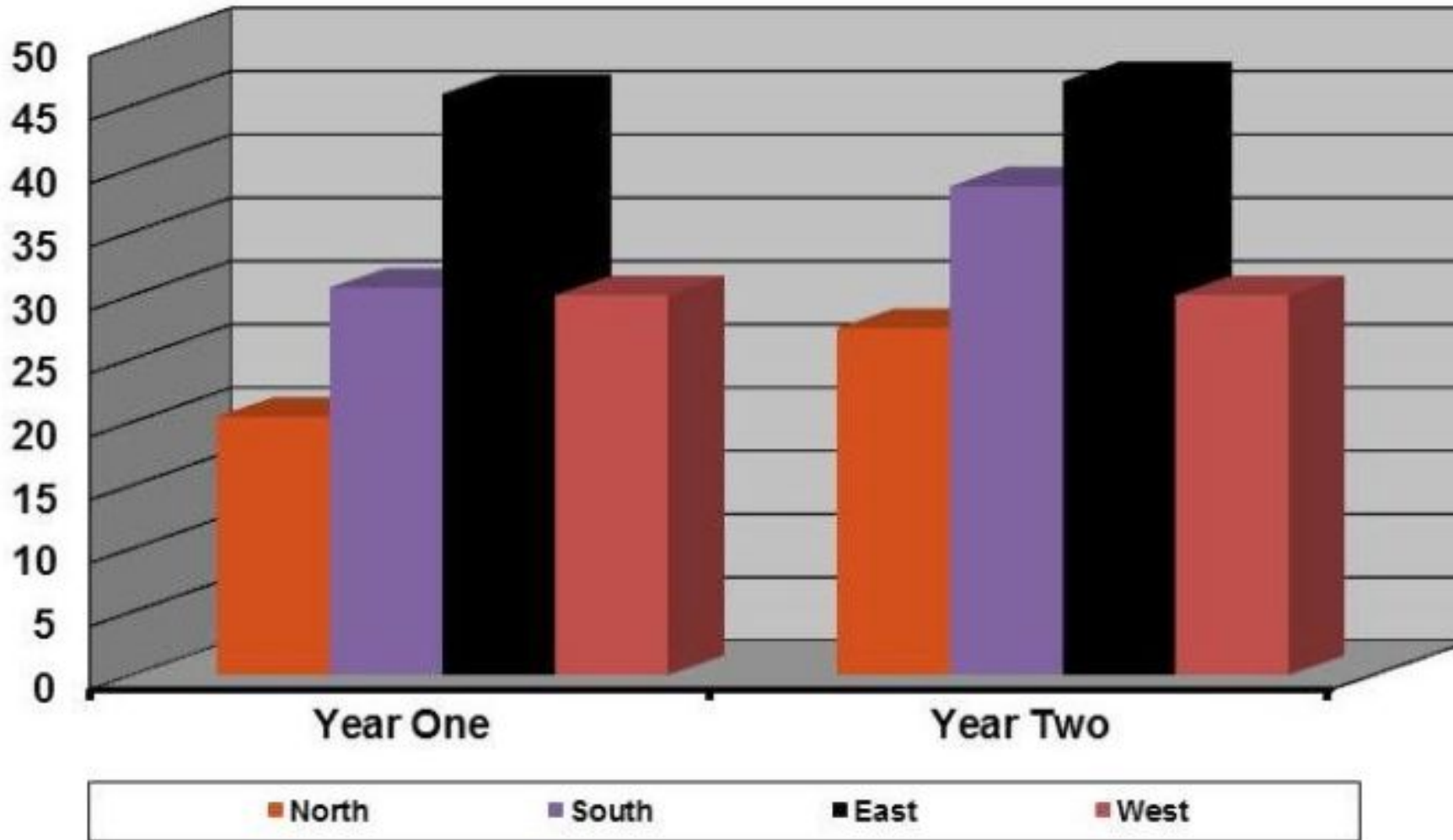


## Investment Portfolio Breakdown



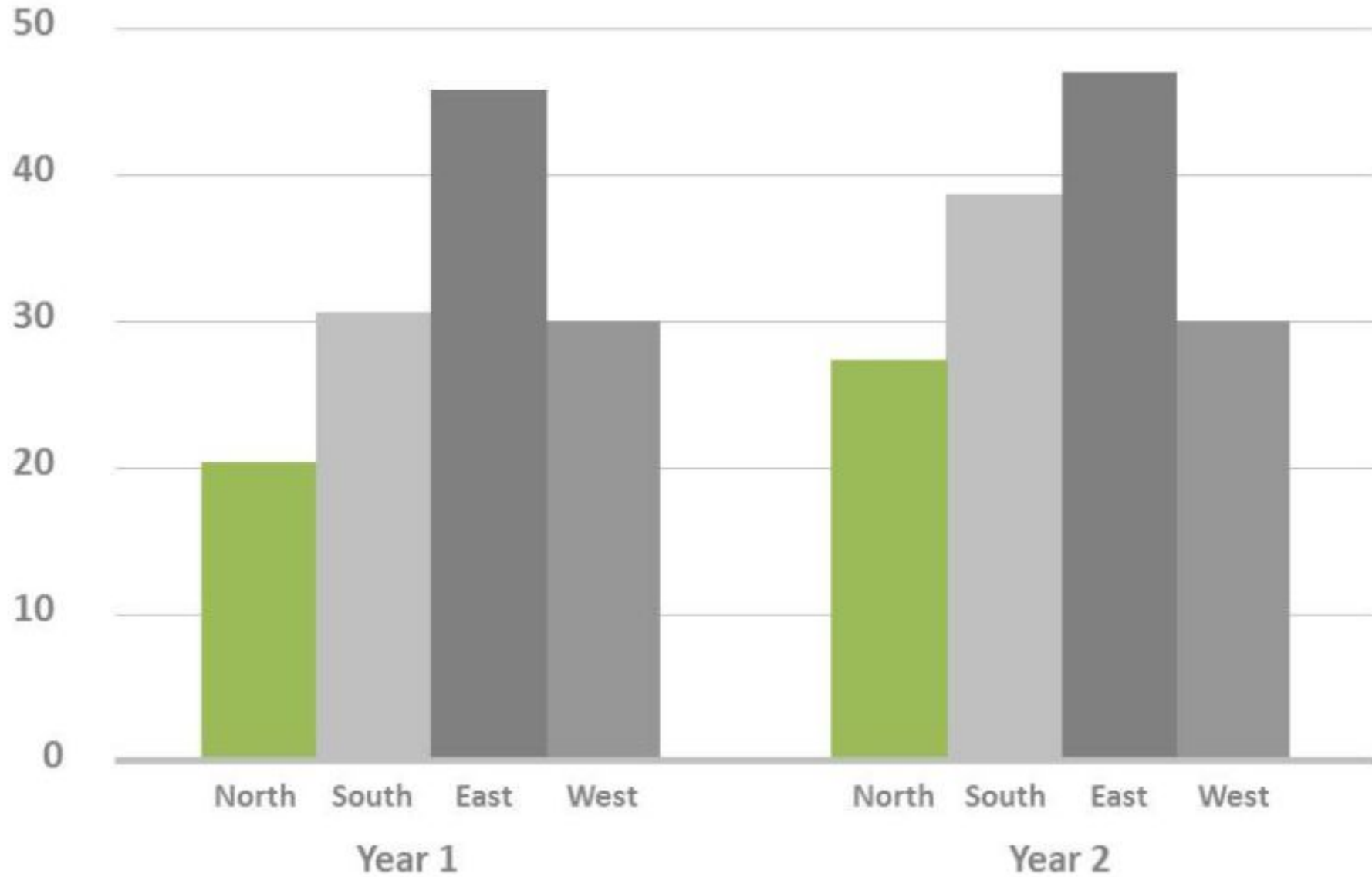
FACULTY OF MANAGEMENT  
School of Information Management

Yearly Revenue Per Region





Yearly Revenue Per Region



**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

# What kind of data do you have?

- Specialized (e.g. geospatial) – outside the scope of this course, but there are many specialized resources
- General quantitative data types
  - Nominal (discrete)
  - Ordinal (greater or less than)
  - Interval-Ratio (continuous)
    - Ratio has a meaningful zero

# Nominal

- Obtained when names or labels are used to (often arbitrarily) identify attributes
- Indicates that there is a difference – classification, e.g.,
  - Male/Female
  - BA, BSc, MA, PhD, JD, MLIS, MBA, MSc, Mphil
  - NS, NB, NL, ON, QC, BC...

# Nominal Data

- Can use numbers
  - But they don't have their traditional 'valued' meaning
  - e.g. USPC technology classifications: 474, 470, 270
- Can't take the average of nominal data
- Can count them, and do cross-tabs
- Graph - Bar chart and related graphs

# Ordinal Data

- How would you rate these slides?
  - Poor
  - Fair
  - Average
  - Good
  - Most Excellent
- Contains information about what is better/worse, but not about the degree

# Ordinal Data

- Allows for ordering/ranking
- Can't take the mean, because the values don't have meaning
- But you can take the mode or median
- Bar charts are probably your best bet again

# Somewhere in between

- The Instructor showed genuine concern for my learning:
  - 1 = Strongly disagree
  - 5 = Strongly agree
- People report the mean of this sort of measurement (Likert scale) frequently
- Can be problematic though

# Interval-Ratio Data

- Ranking and precise distance between values can be measured
  - Unlike ordinal that just ranks data and doesn't allow for differences to be measured
- E.g.,
  - Age
  - Temperature
  - Income
  - Cost

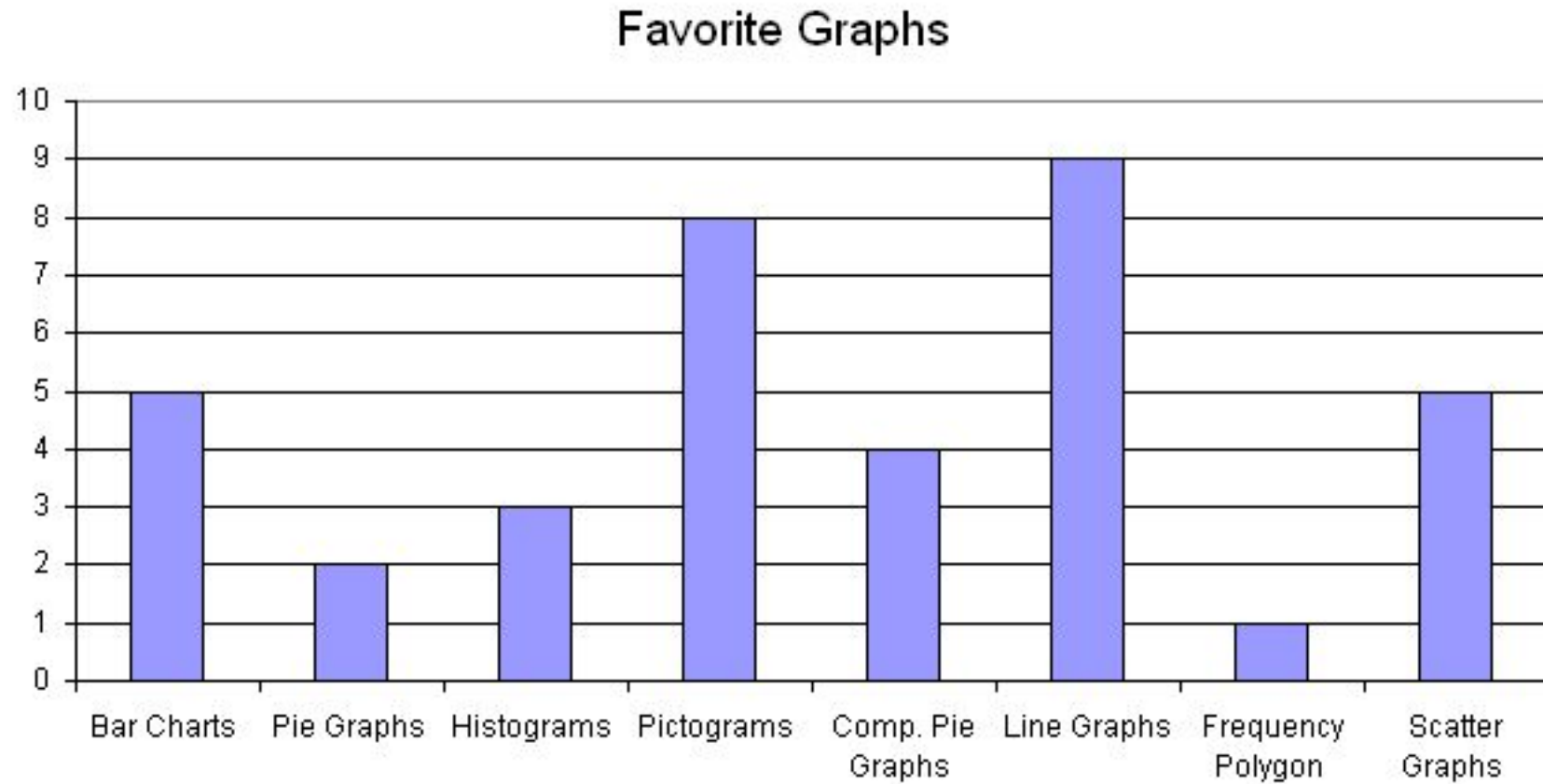


# Interval/Ratio Data

- Interval variables have no true zero
  - Temperature (in practice, but not in theory)
- Ratio has a true zero
  - Income
  - Weight
- Graphs: Line graph, area chart

# Types of graphs/charts

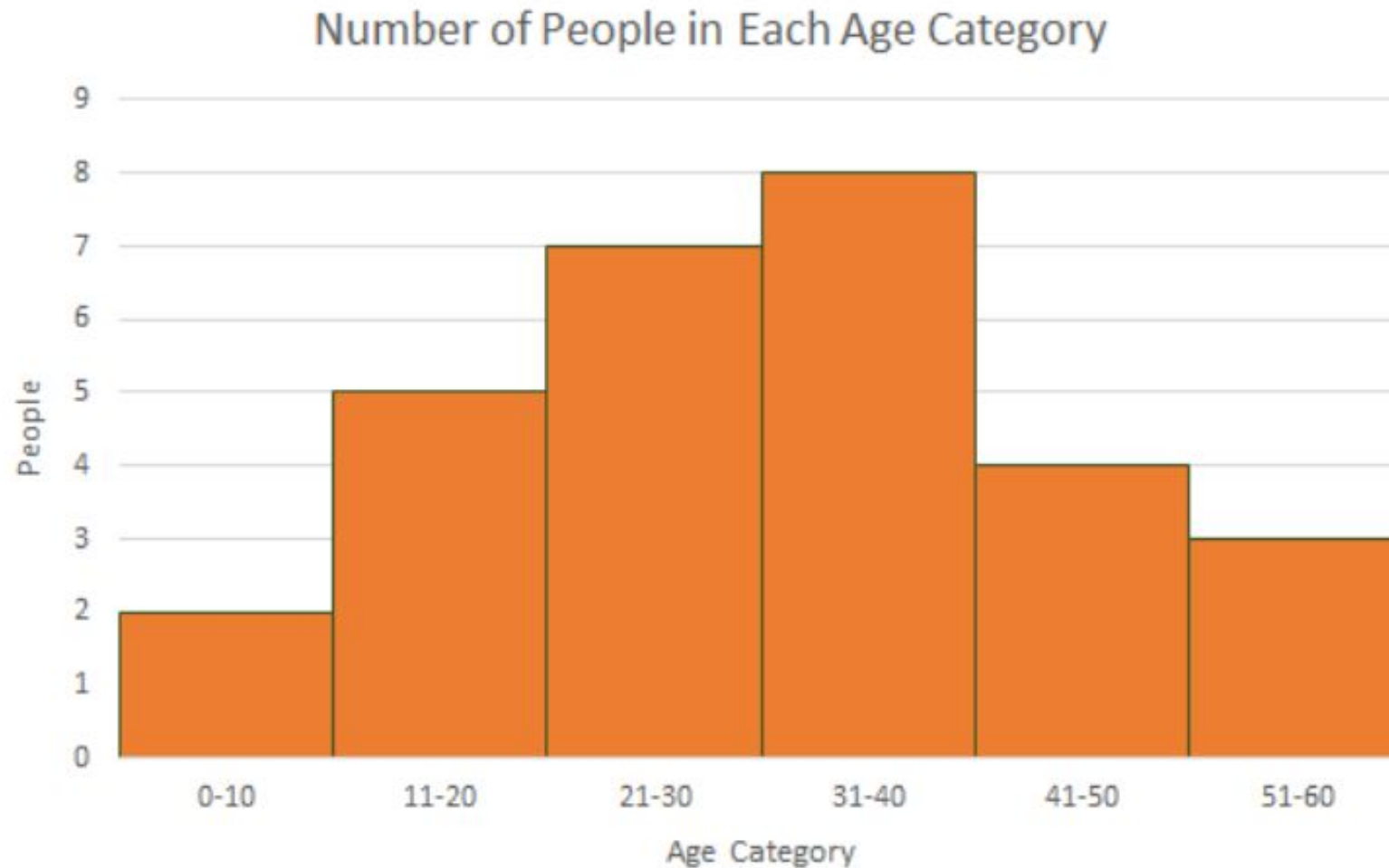
# Bar chart



**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

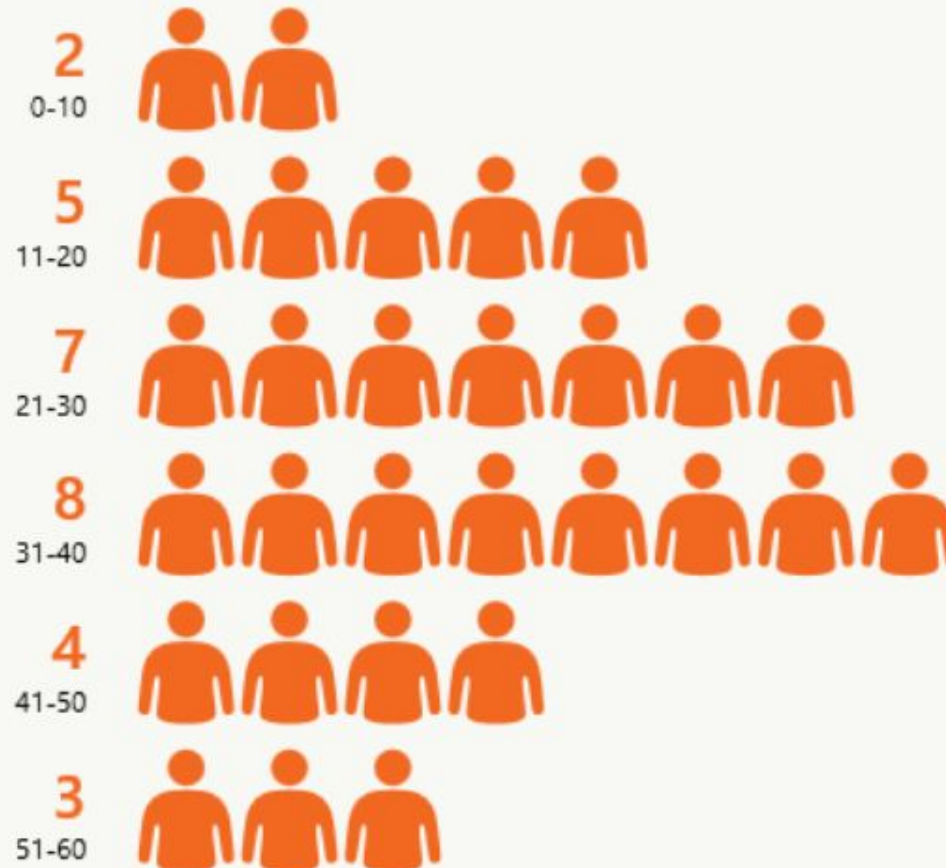
# Histogram



# Pictogram

## Number of People in Each Age Category

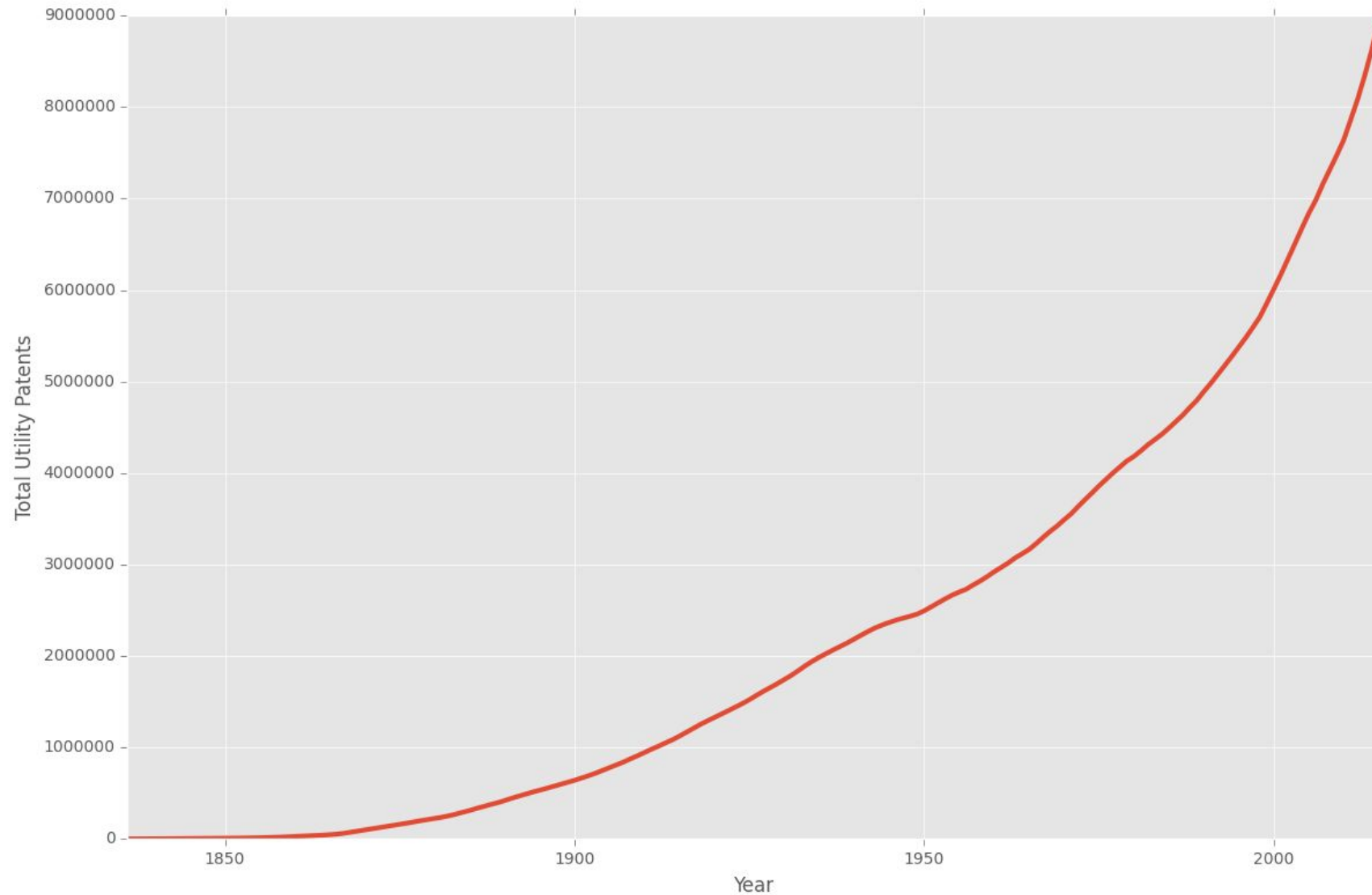
People  
Age Category



**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

# Line graph

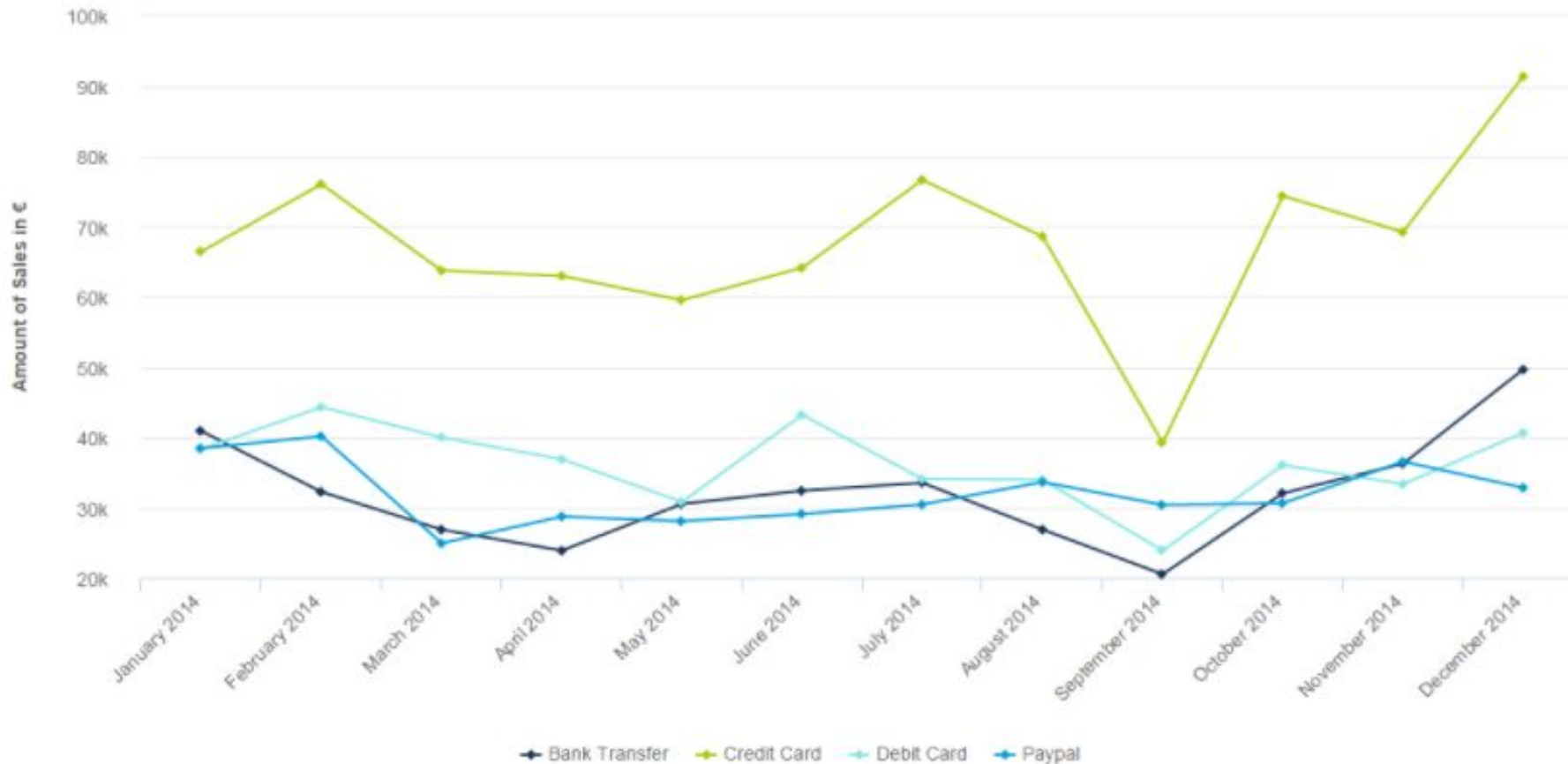


**DALHOUSIE  
UNIVERSITY**

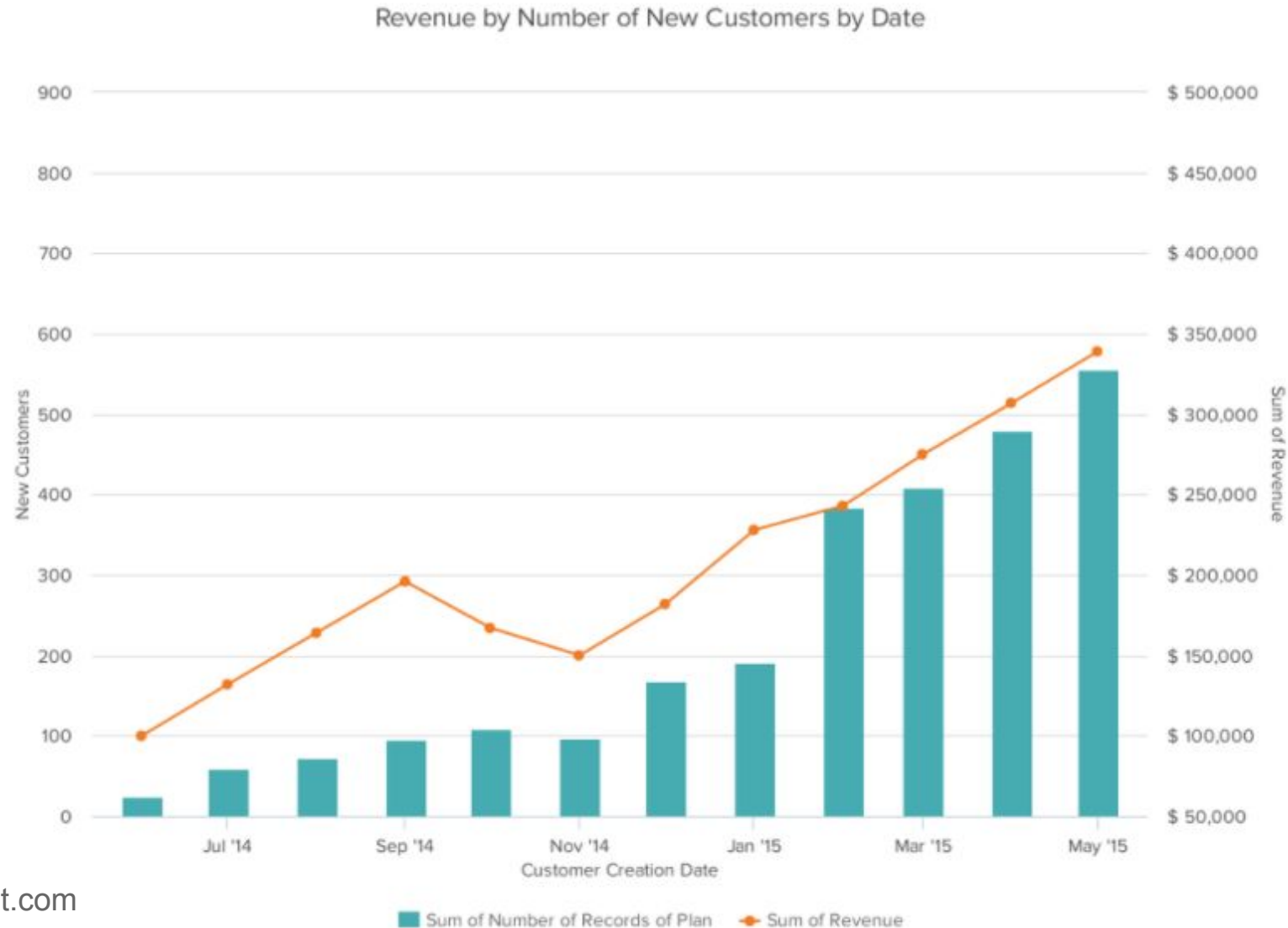
FACULTY OF MANAGEMENT  
School of Information Management

# Line graph

Amount of Sales by Payment Method (last year)

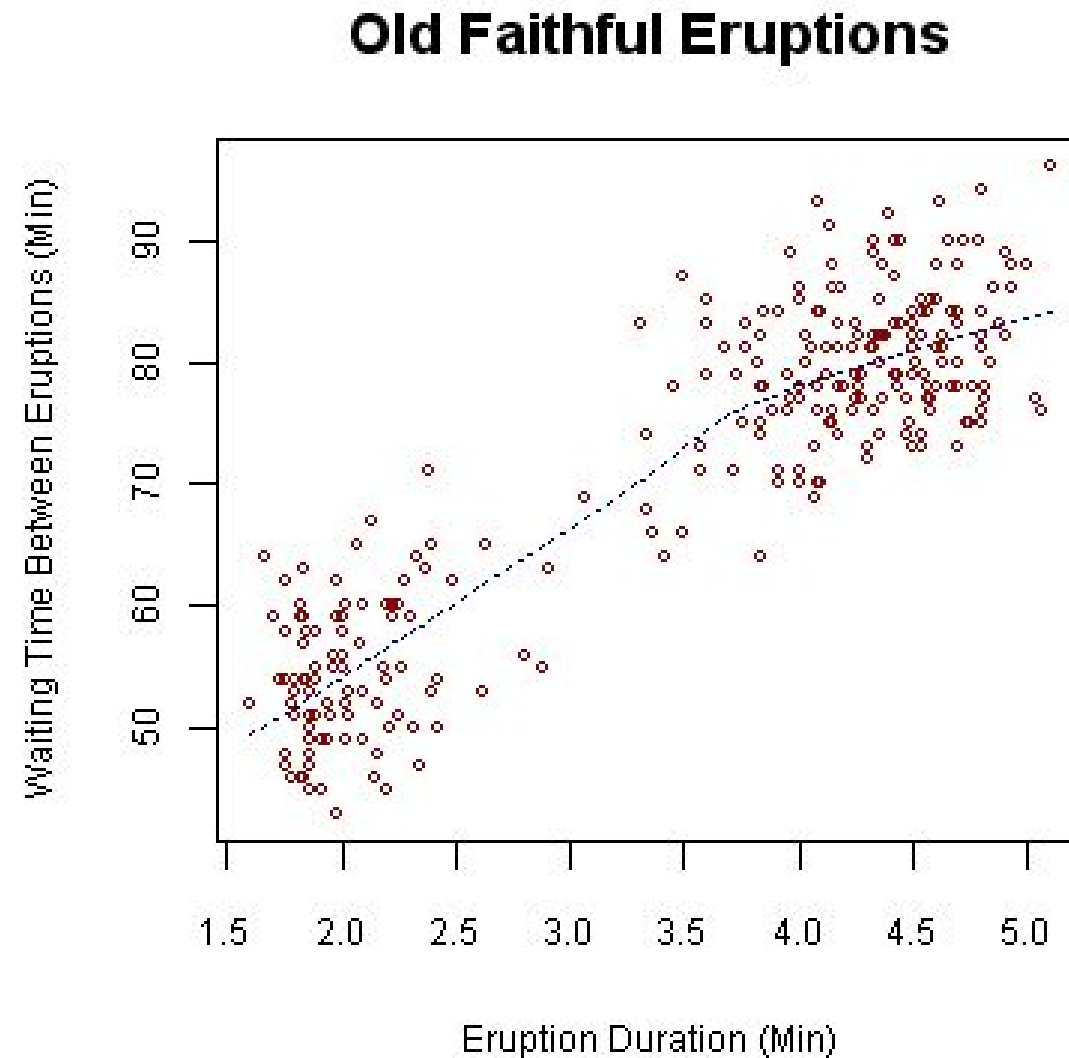


# Dual axis

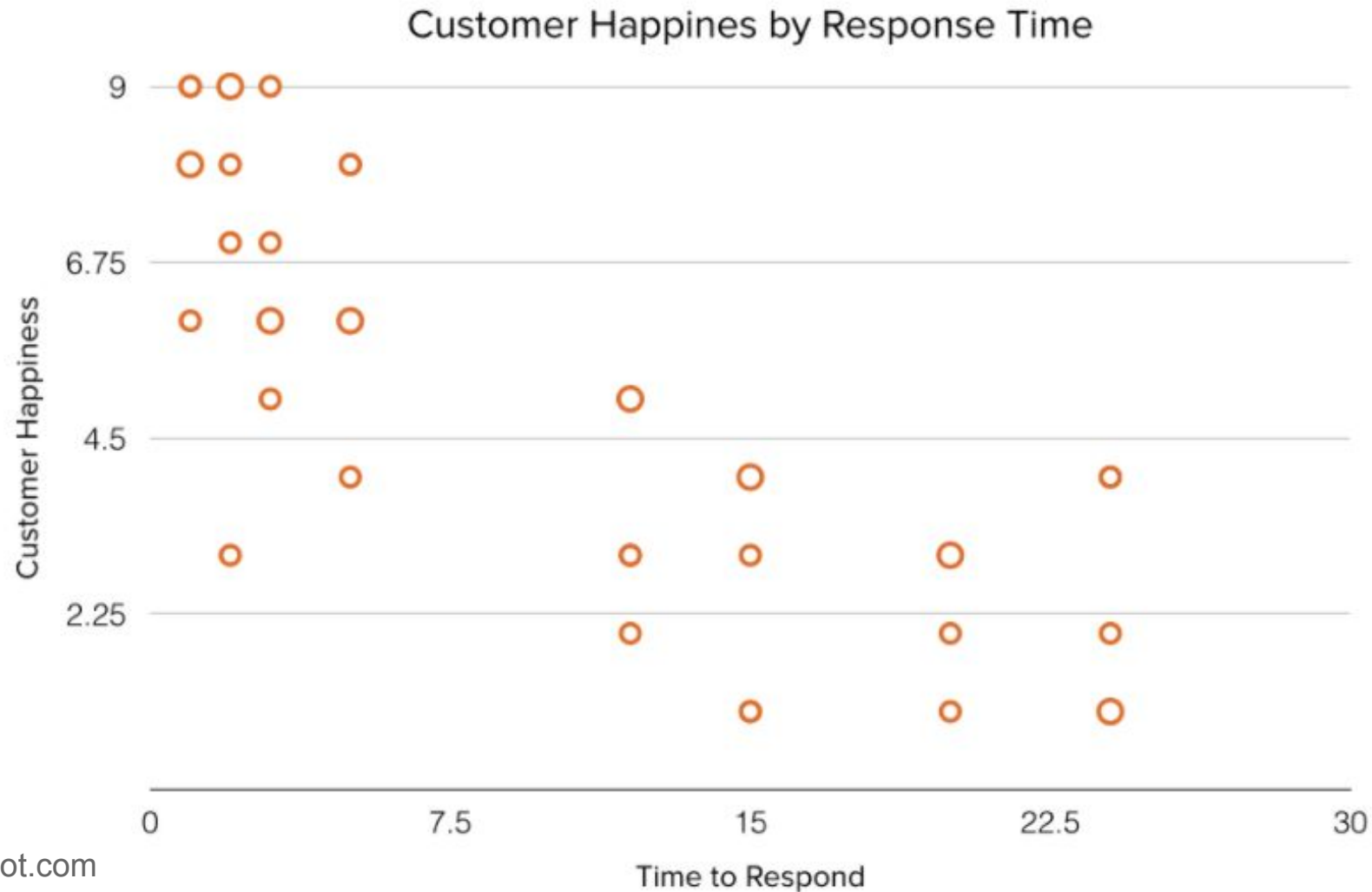




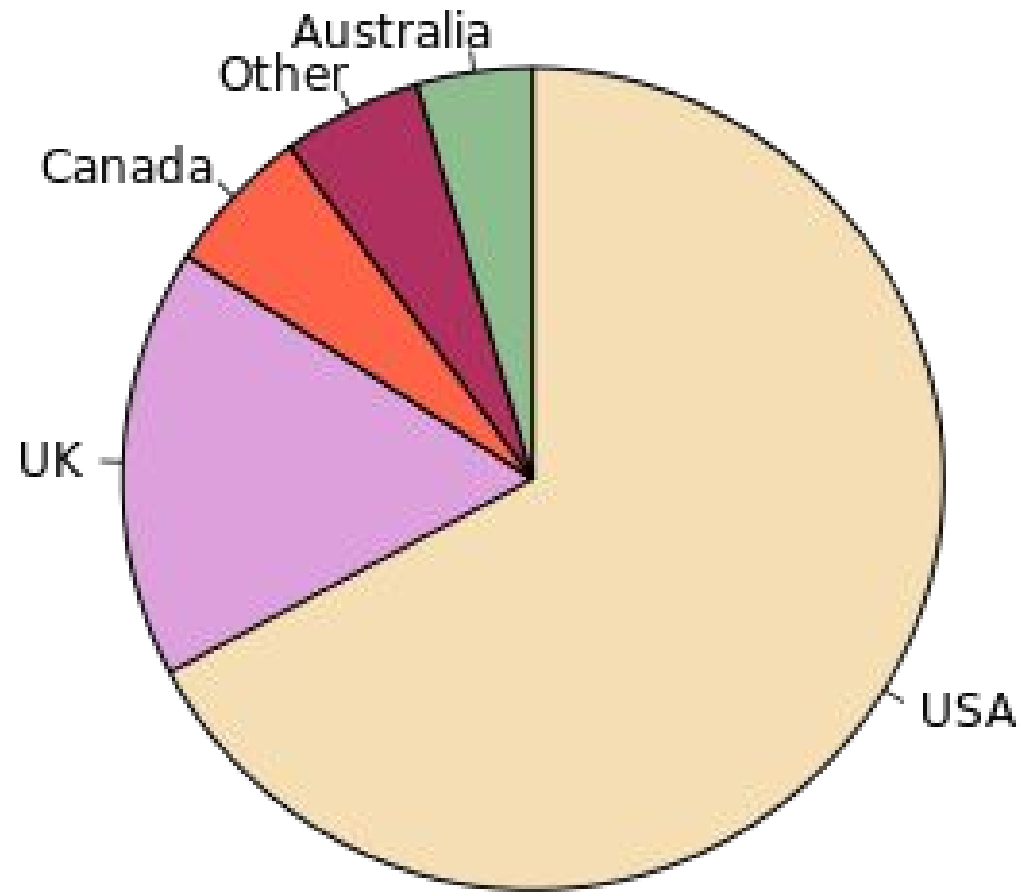
# Scatter plot



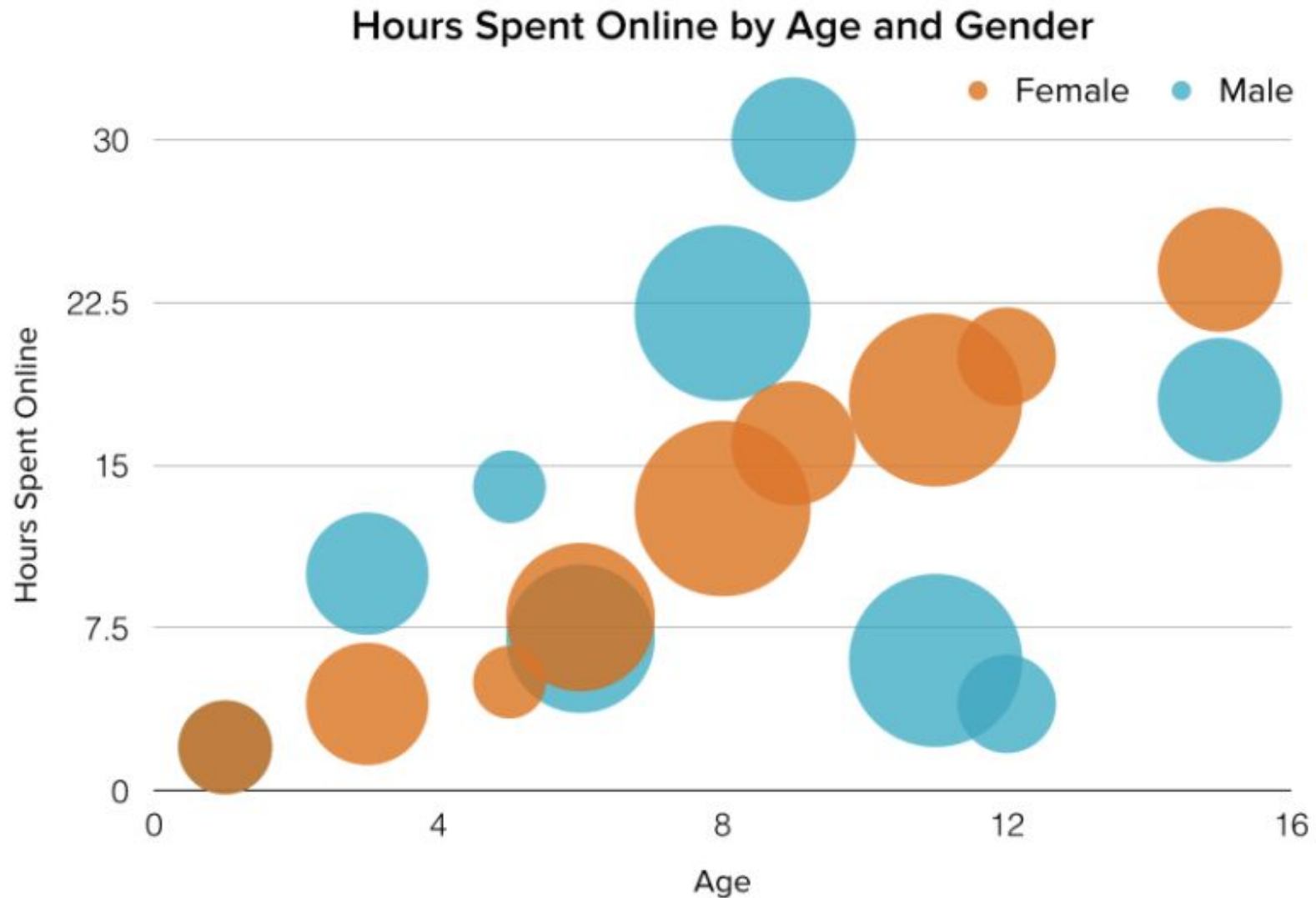
# Scatterplot



# Pie chart



# Bubble chart



# To summarize

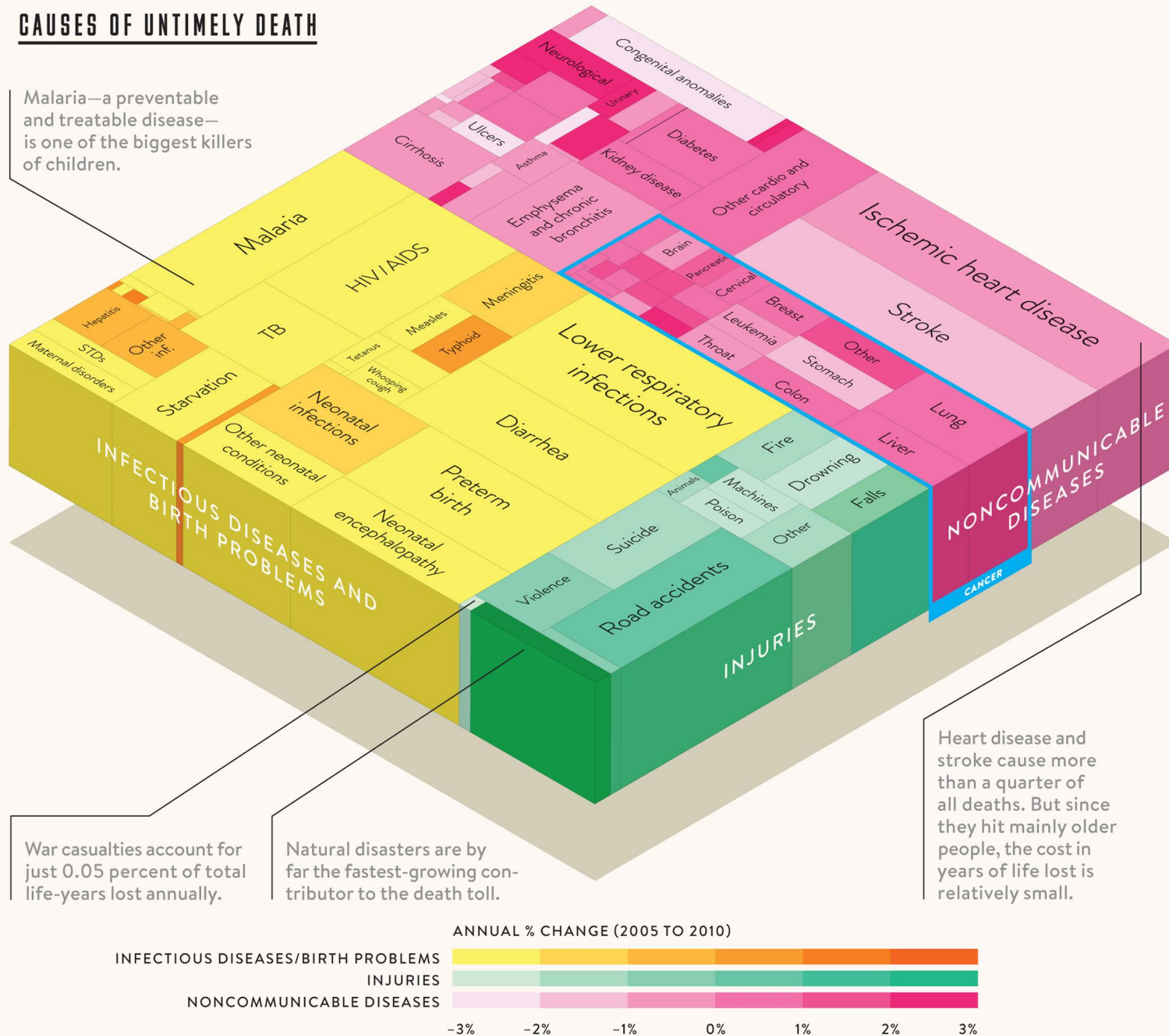


**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

## CAUSES OF UNTIMELY DEATH

Malaria—a preventable and treatable disease—is one of the biggest killers of children.



War casualties account for just 0.05 percent of total life-years lost annually.

Natural disasters are by far the fastest-growing contributor to the death toll.

Heart disease and stroke cause more than a quarter of all deaths. But since they hit mainly older people, the cost in years of life lost is relatively small.

## Global causes of lost life



FACULTY OF MANAGEMENT  
School of Information Management

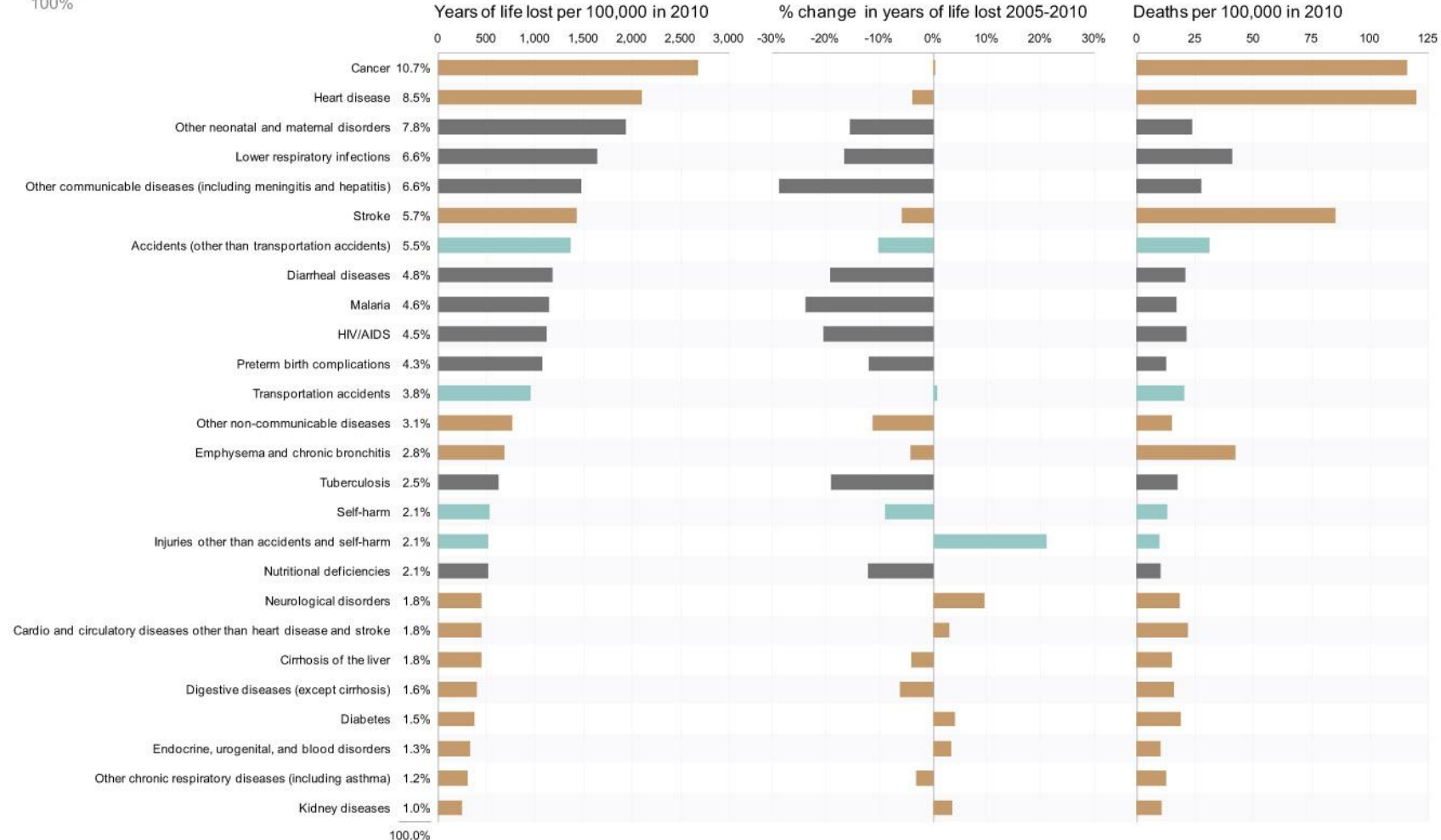
“I love this graph because it shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. In fact, fewer kids are dying, more kids are going to school and more diseases are on their way to being eliminated. But there remains much to do to cut down the deaths in that yellow block even more dramatically. We have the solutions. But we need to keep up the support where they're being deployed, and pressure to get them into places where they're desperately needed.”

~ Bill Gates, WIRED

# Global Causes of Lost Life

44% ■ Communicable, maternal, neonatal, and nutritional disorders  
43% ■ Non-communicable diseases  
13% ■ Injuries  
100%

Comparing the number of deaths alone, as shown in the right-most graph below, doesn't tell the entire story. Some causes of death have a greater effect on the young, which can be seen when comparing years of life lost in the leftmost graph.



Some causes of death contribute disproportionately to years of life lost because of their effect on the young. For example, malaria, while not huge in the number of deaths, is much more significant in the number of years that are lost.

Two interesting changes reside in "Injuries other than accidents and self-harm." War, which accounted for only 0.05% of years of life lost, decreased since 2005 by 31.5% in years of life lost per 100,000 people. Natural disasters, which accounted for 0.65% of years of life lost, increased by 217% in years of life lost per 100,000.

Communicable, maternal, neonatal, and nutritional disorders (the gray bars) are often easier to prevent through healthcare than other causes of death. This reveals itself in the graph above by the fact that all of these disorders have decreased during this five year period.

The five forms of cancer that cause the most deaths are trachea/bronchus/lung (2.9%), stomach (1.4%), liver (1.4%), colon/rectum (1.4%), and breast (0.8%).

All cardiovascular and circulatory diseases combined account for 30% of deaths.



**DALHOUSIE  
UNIVERSITY**

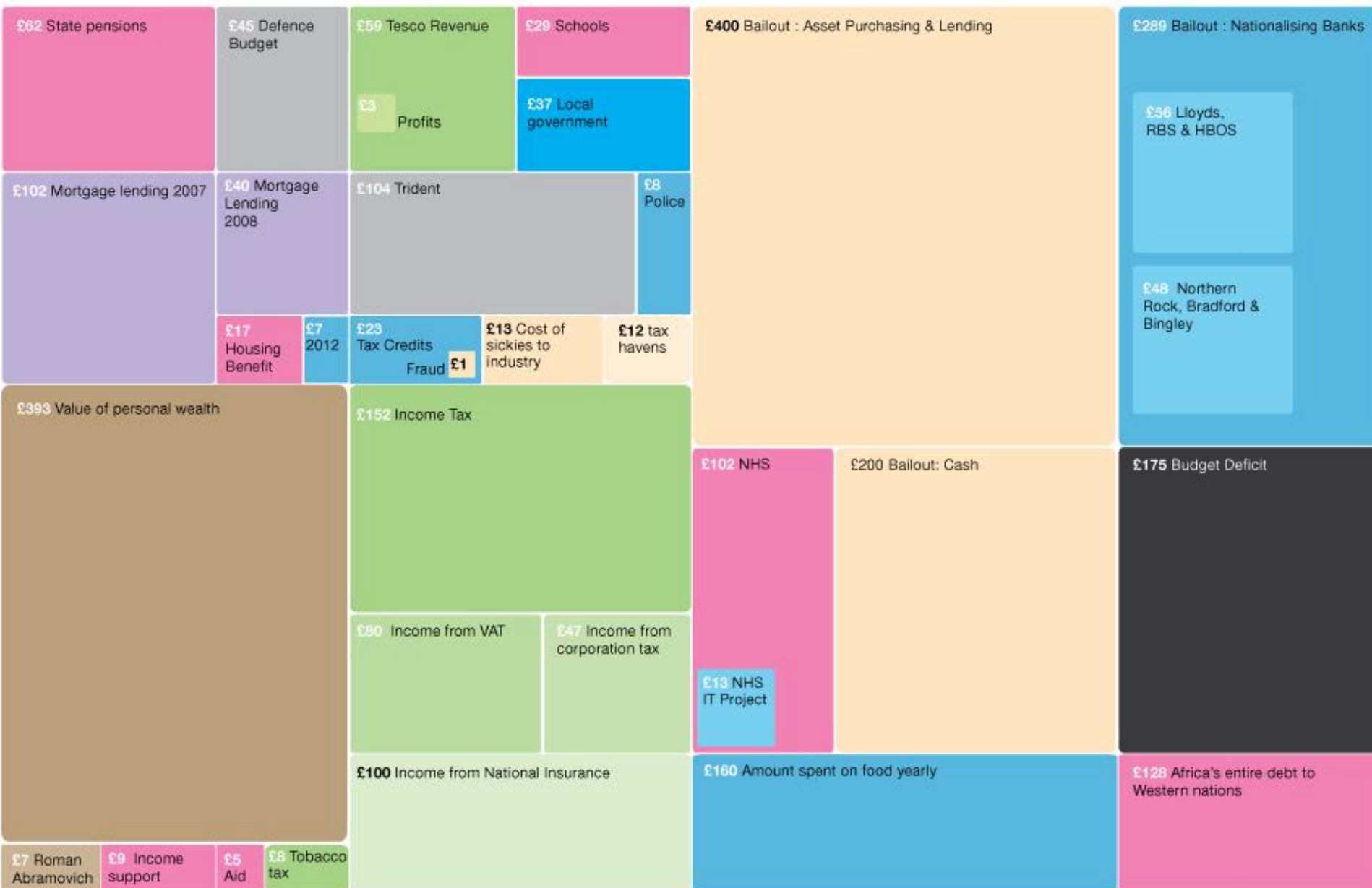
FACULTY OF MANAGEMENT  
School of Information Management



# British budget deficit



FACULTY OF MANAGEMENT  
School of Information Management



Giving Spending Fighting Hoarding Lending Bailing Earning

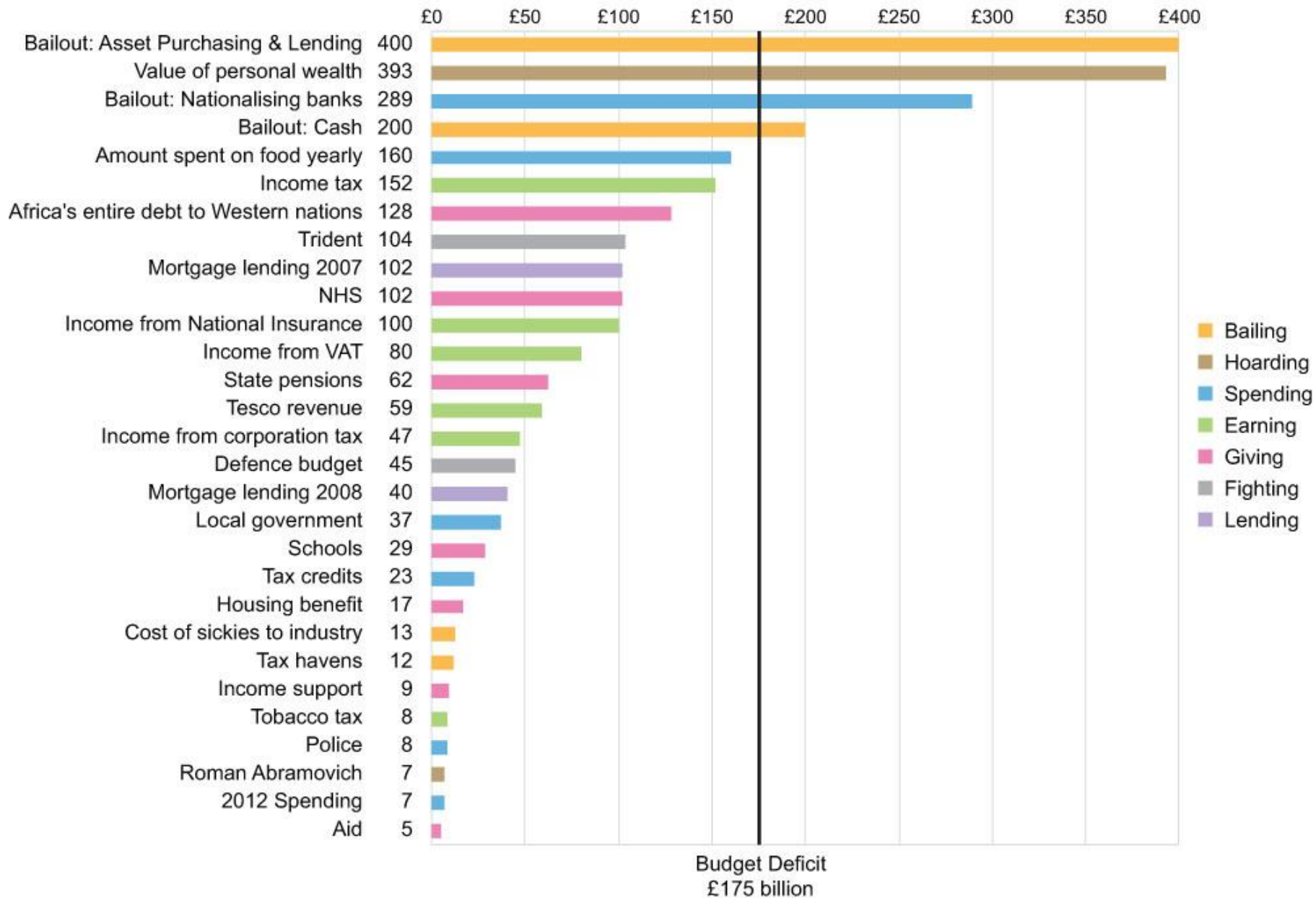
The Billion Pound-O-Gram

David McCandless / InformationIsBeautiful.net

Source: UK Treasury, Guardian

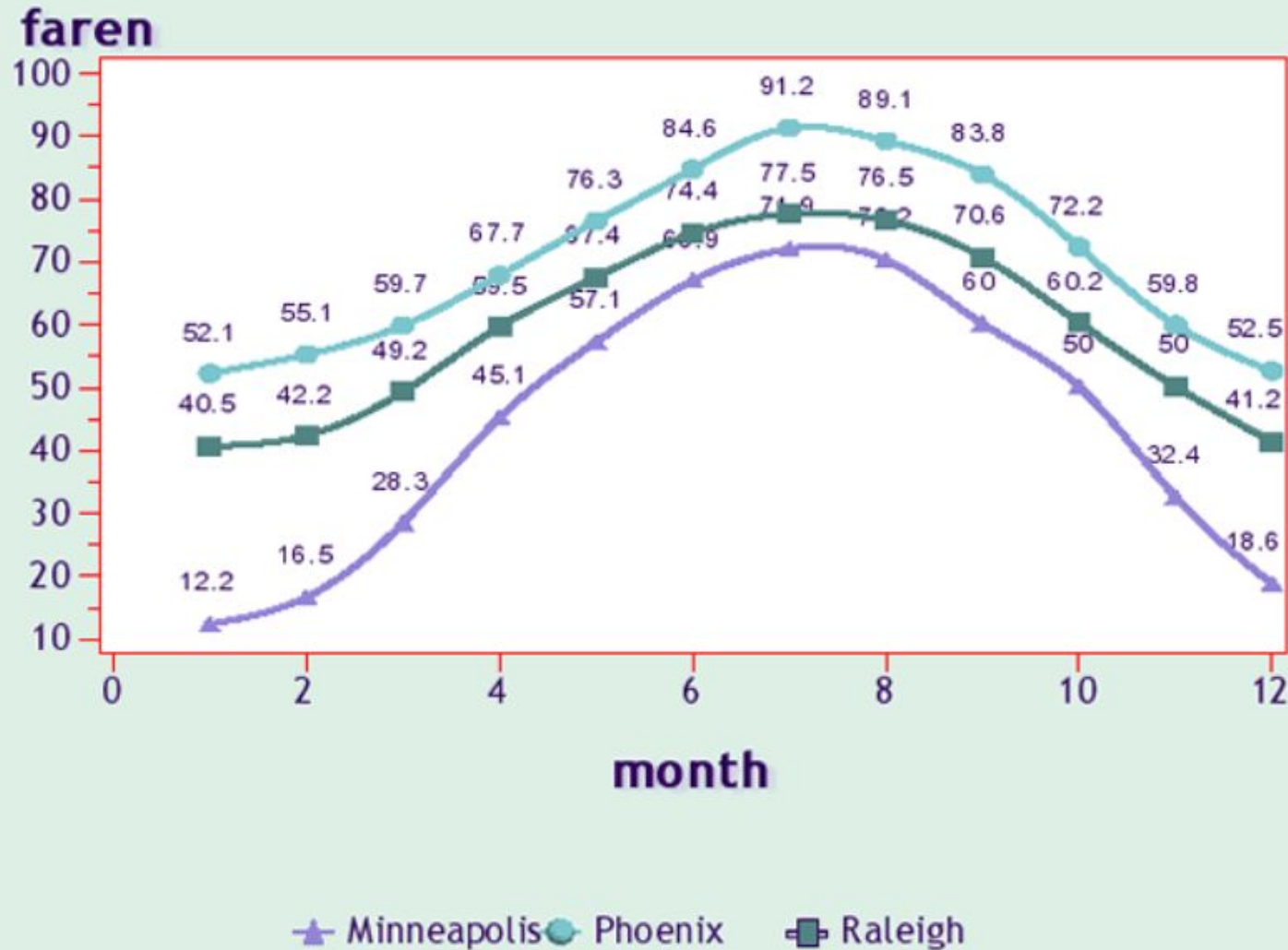
# The UK Budget Deficit Compared to Other Big Amounts of Money

(Billions of Pounds)



FACULTY OF MANAGEMENT  
School of Information Management

# Average Monthly Temperature

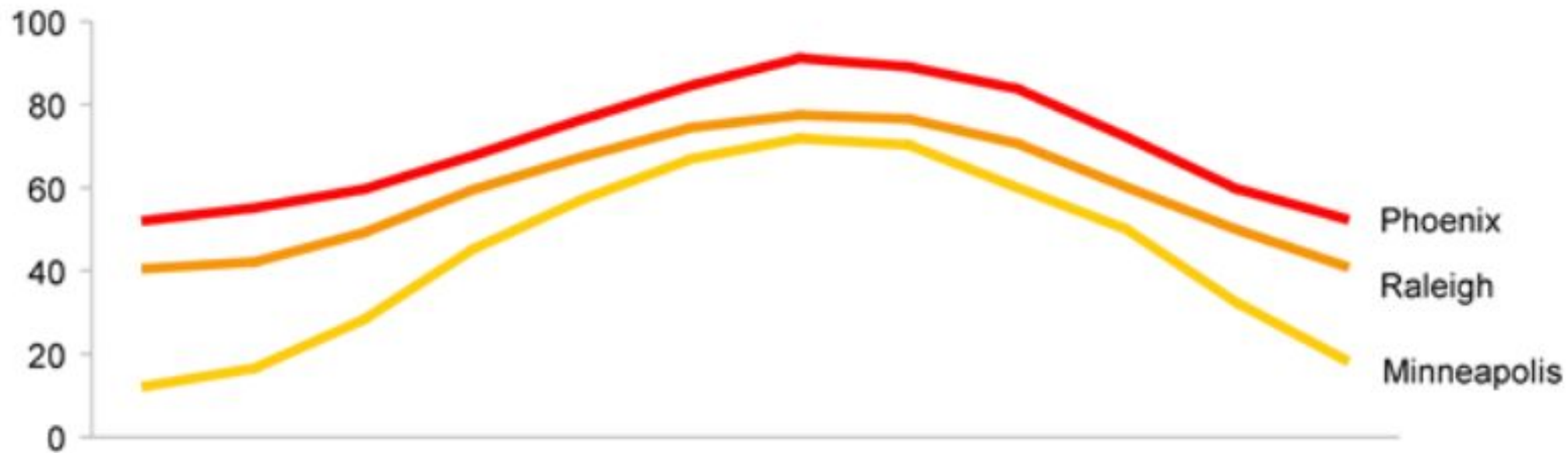


**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

Fahrenheit

### Average (Mean) Monthly Temperatures in 2003



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Phoenix	52.1	55.1	59.7	67.7	76.3	84.6	91.2	89.1	83.8	72.2	59.8	52.5
Raleigh	40.5	42.2	49.2	59.5	67.4	74.4	77.5	76.5	70.6	60.2	50.0	41.2
Minneapolis	12.2	16.5	28.3	45.1	57.1	66.9	71.9	70.2	60.0	50.0	32.4	18.6



**DALHOUSIE  
UNIVERSITY**

FACULTY OF MANAGEMENT  
School of Information Management

# When designing, keep in mind

- What information is needed?
- What do colors mean?
- Which icons are familiar to the audience
- Use a descriptive, yet brief, headline
- Use the right type of chart
- Size and position of elements are important
- K.I.S.S. (Keep It Simple Silly)

# Gestalt laws of grouping

[https://en.wikipedia.org/wiki/Principles\\_of\\_grouping](https://en.wikipedia.org/wiki/Principles_of_grouping)

[www.plot.ly](http://www.plot.ly)