SOC40830 Quantitative Data Analytics and Applications

Dr. Eoin Flaherty (D413, Newman Building)
eoin.flaherty@ucd.ie

Week 4, Monday October 3rd

1

Week 3 Outline

1. **Online Data Visualization**
2. **Historical Data Visualization**
3. **Key Principles**
4. **Graphing in Stata I**

2

1. Online Data Visualization

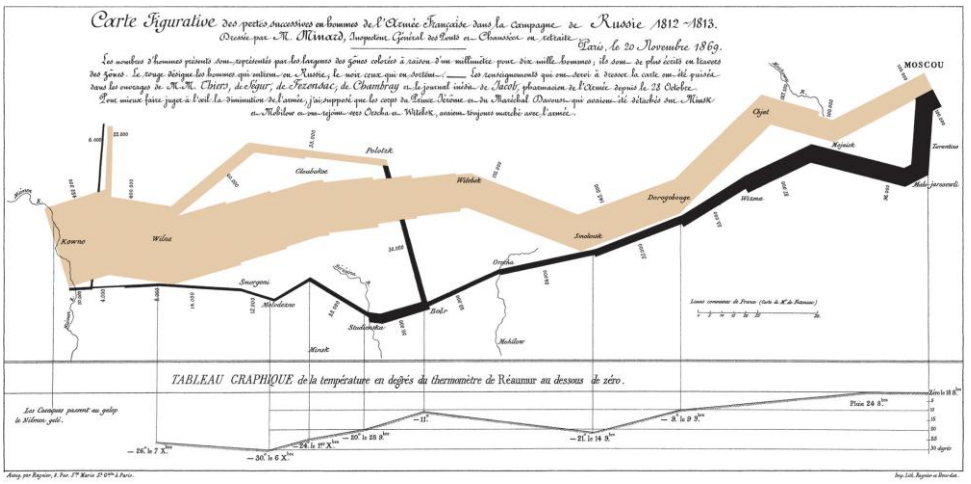
www.gapminder.org

www.airo.ie

3

2. Historical Data Visualization

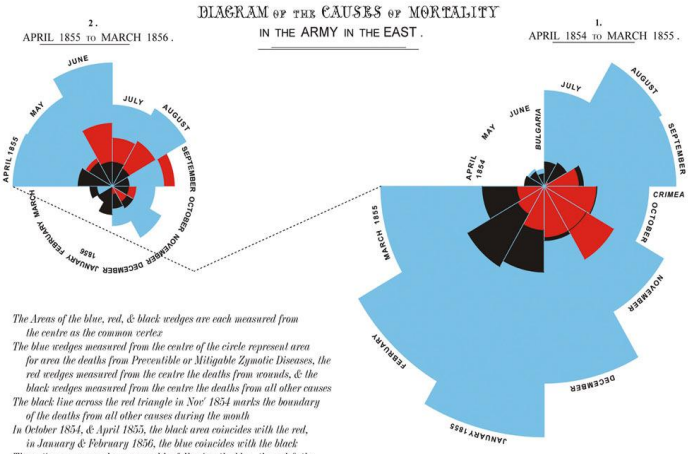
Minard's Map Of Napoleon's March on Russia



4

2. Historical Data Visualization

Nightingale Roses



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.couk

5

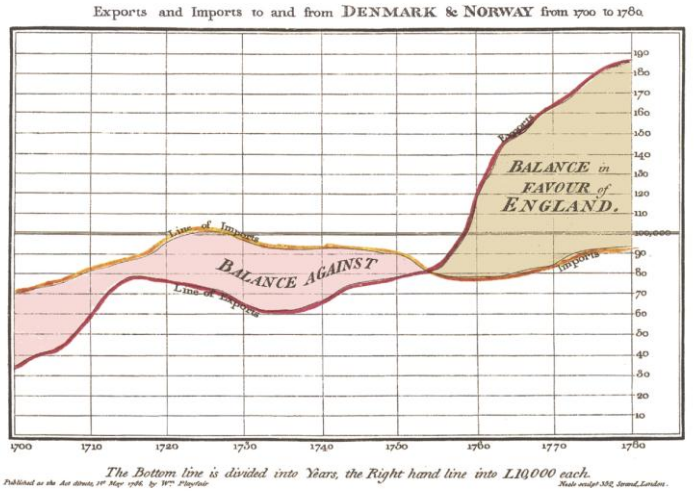
2. Historical Data Visualization

John Snow (1854)

6

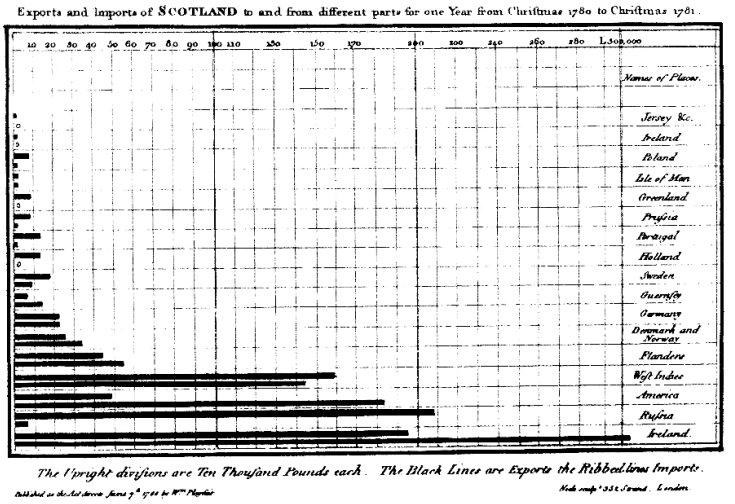
2. Historical Data Visualization

William Playfair (1786)



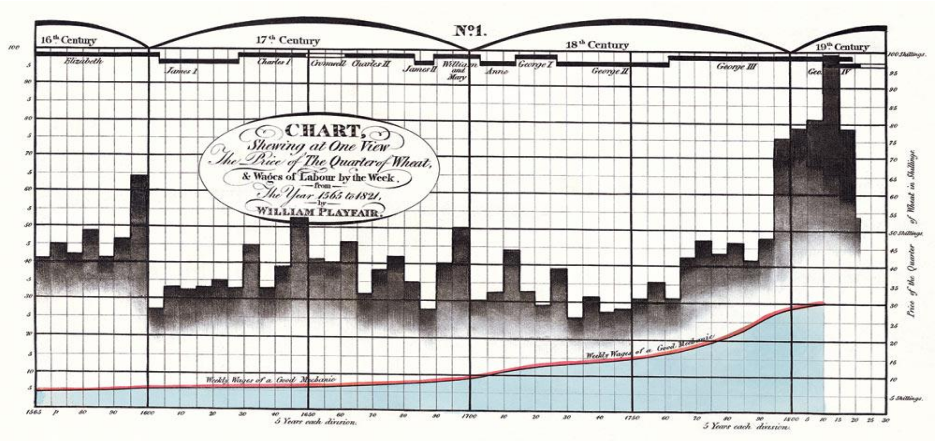
2. Historical Data Visualization

William Playfair (1786)



2. Historical Data Visualization

William Playfair (1786)



9

3. Key Principles

Why Graph?

1. Show data without distortion, complex representation in small space, overcome limitations of displaying large volumes of data.

2. Reveal patterns, integrate verbal and statistical description, exploration, multilevel representation – fine to macro-detail.

3. Comparison and inspiration, comparing cases and sub-groups, independent/dependent association.

10

3. Key Principles

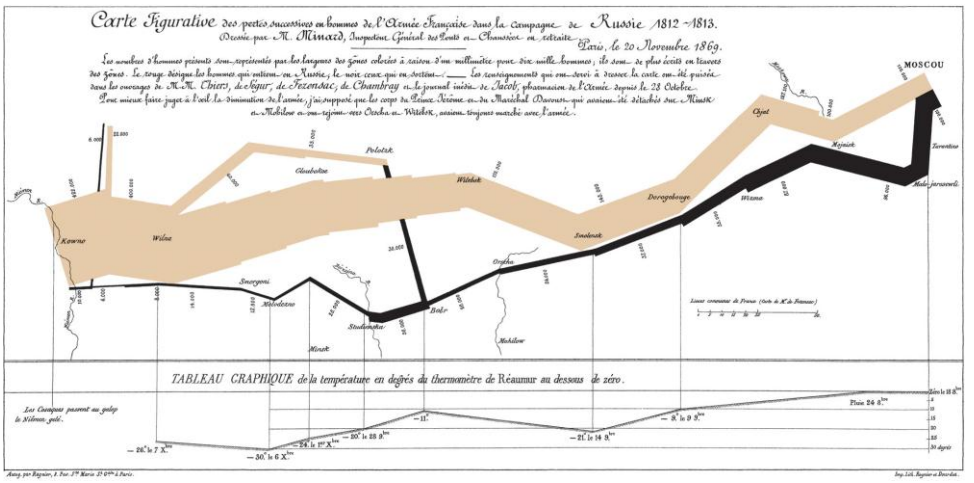
Tufte (1983)

Graphical excellence…

- …consists of complex ideas communicated with clarity, precision and efficiency."
- …is well designed presentations of interesting data – a matter of *substance*, of *statistics* and of *design*.
- …gives the reader the greatest number of ideas in the shortest time with the least ink in the smallest space.
- …requires telling truth about data.

11

3. Key Principles



12

3. Key Principles

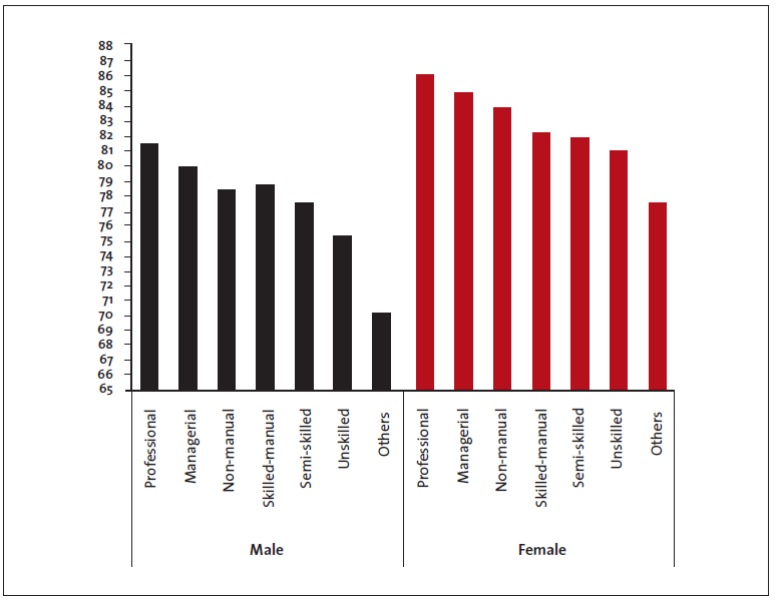
How to Distort

- Occurs when the visual representation of the data is not consistent with the numerical representation

- Achieved through:
  - The use of area
  - Manipulation of axis
  - Chart junk (and other unnecessary information)
  - 3D effects

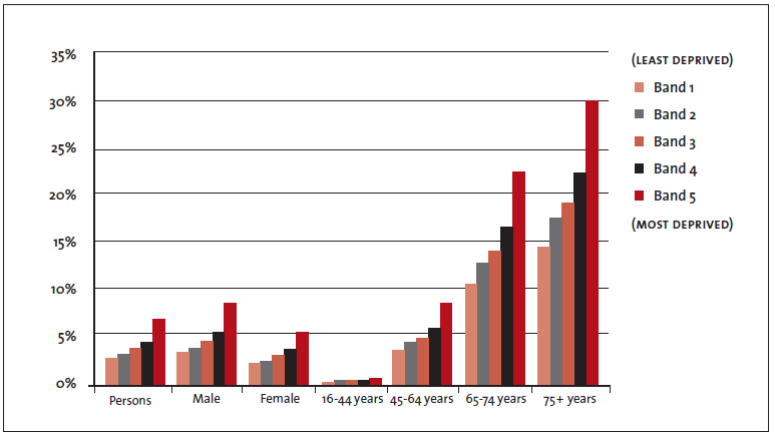
13

Figure 3: Life expectancy at birth by social class



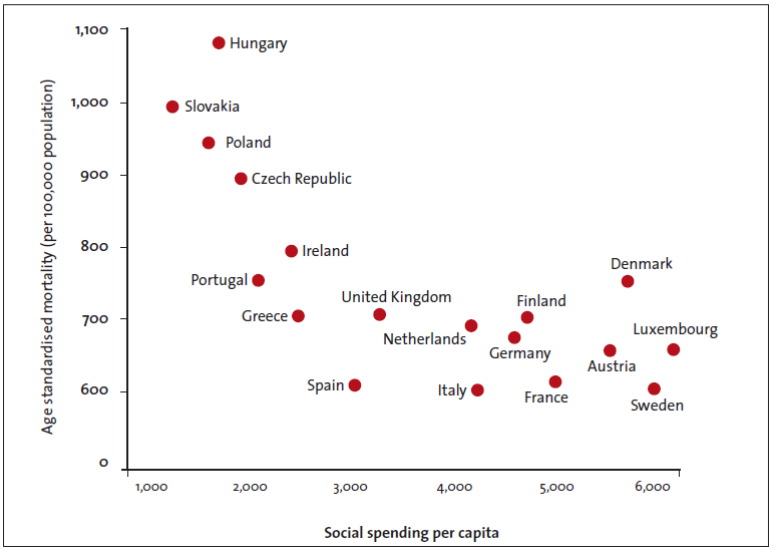
14

Figure 4: Population prevalence rates of angina and heart attack (CHD) amongst adults across the deprivation bands in the Republic of Ireland within each sex and age group (2007)[13]



15

Figure 7: Relation between social welfare spending and all-cause mortality in 18 EU countries, 2000



16

3. Key Principles

SMR = (Observed Deaths / Expected Deaths) x 100

	Individuals	Observed Deaths	Death rate	Expected Deaths
Population	434,217	512	.12 per 100	-
Group A	52,004	47	.09 per 100	
Group B	87,323	140	.16 per 100	

17

3. Key Principles

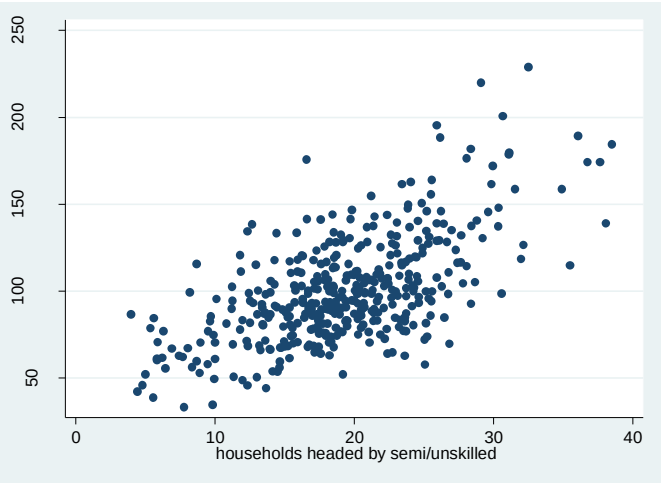
SMR = (Observed Deaths / Expected Deaths) x 100

	Individuals	Observed Deaths	Death rate	Expected Deaths
Population	434,217	512	.12 per 100	-
Group A	52,004	47	.09 per 100	62
Group B	87,323	140	.16 per 100	105

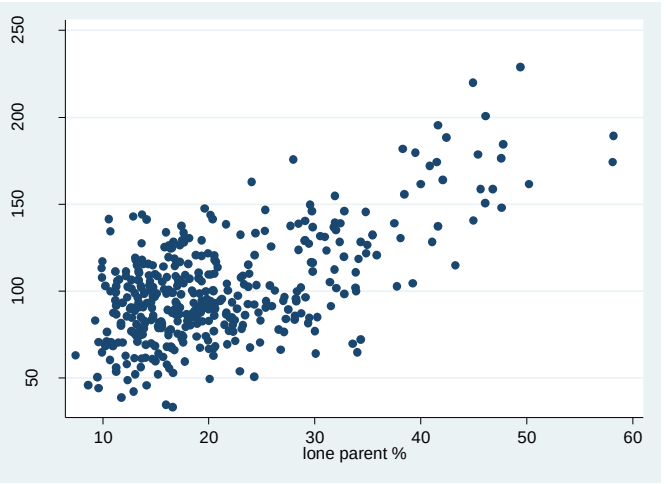
SMR (Group A) = (47/62)*100 = **75.8**

SMR (Group B) = (140/105)*100 = **133.3**

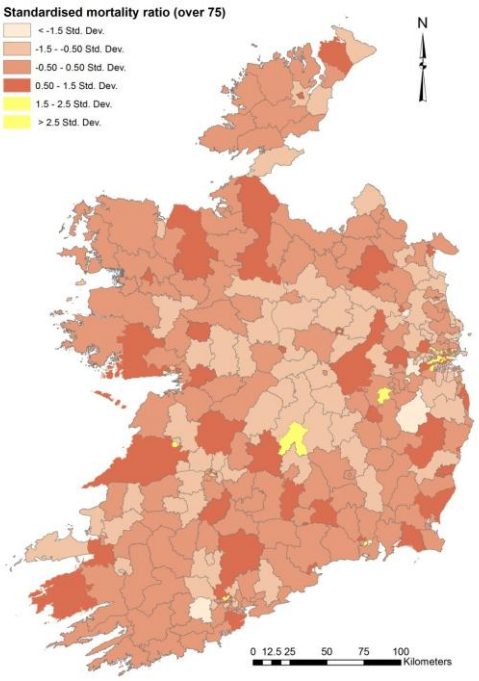
18


250
200
150
100
50
0
10
20
30
40
households headed by semi/unskilled

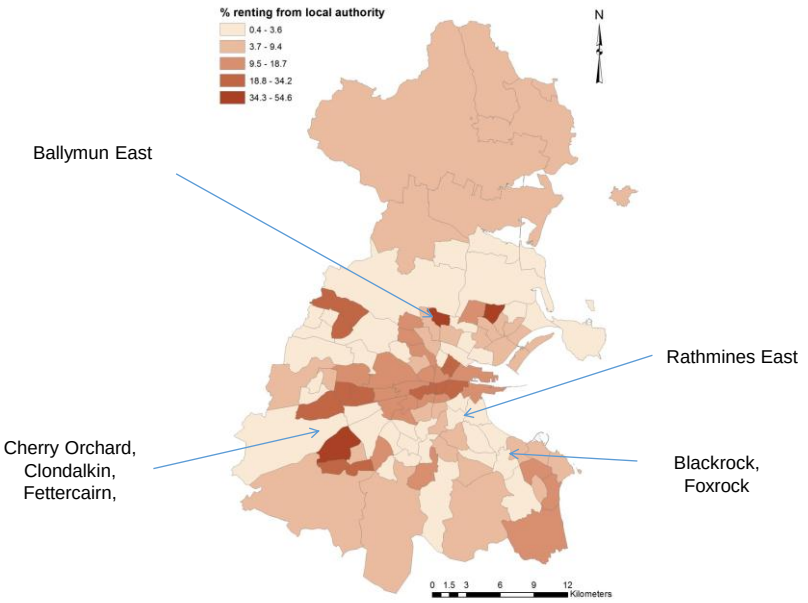
19


250
200
150
100
50
10
20
30
40
50
60
lone parent %

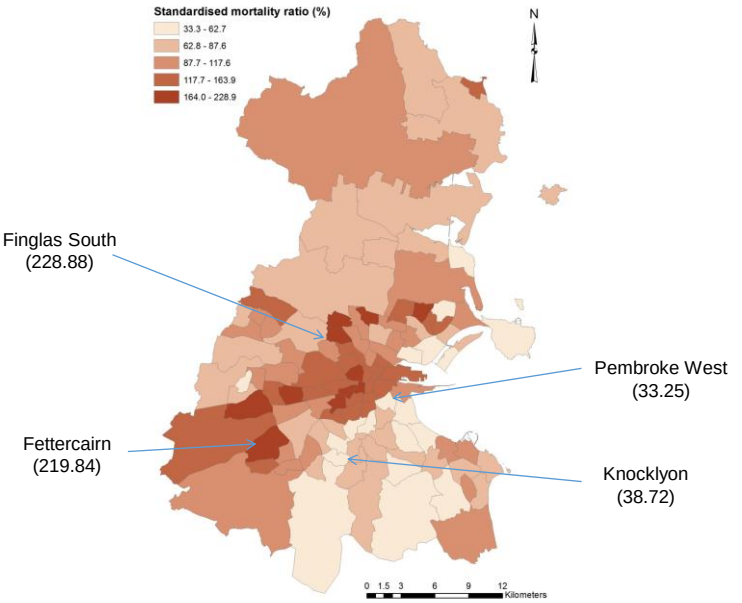
20


Standardised mortality ratio (over 75)
< -1.5 Std. Dev.
-1.5 - -0.50 Std. Dev.
-0.50 - 0.50 Std. Dev.
0.50 - 1.5 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
N
0 12.5 25 50 75 100
Kilometers

21


% renting from local authority
0.4 - 3.6
3.7 - 9.4
9.5 - 18.7
18.8 - 34.2
34.3 - 54.6
N
Ballymun East
Rathmines East
Cherry Orchard,
Clondalkin,
Fettercairn,
Blackrock,
Foxrock
0 1.5 3 6 9 12
Kilometers

22



23

4. Graphing in Stata I

Group Exercise

1. Open Stata
2. Download the **international dataset** from Blackboard
3. Explore the dataset using **`des`** and **`sum`**

4. Select two variables from the dataset, choose a visualisation method appropriate to their level of measurement.
5. Describe the distributions using your chosen visualisation method.

24

4. Graphing in Stata I

Group Exercise

1. Open Stata
2. Download the **Irish dataset** from Blackboard
3. Explore the dataset using des and sum

4. Generate time series line graphs of two variables.
5. Describe the dynamics of the data (stability/volatility, key changes).

25