Research Methods

Alex Klein

Why Data Analysis?

- Always a need to provide various types of evidence in support of statements, propositions or conclusions
- Purpose of data analysis is to **transform raw data into usable information**
- Requires ordering and assembling of data into data sets
- Economists frequently rely on secondary data, that is, data collected by someone other than the user
  - Census, surveys, organisational records etc.
- But, increasing use of primary data, that is, data collected by the person undertaking the research
- So data analysis is a key skill of an economist!

Data Analysis & Research

- Economists tend to approach (applied) research in a clear and structured manner:
  - Economic Theory (Hypotheses) →
  - Testable Hypotheses →
  - Data and Measurement Issues →
  - Empirical (Econometric) Methods →
  - Results (interpretation) →
  - Policy Implications
- Challenge then is to write up this research in a **clear, coherent and persuasive manner**

Data Analysis & Research

- Common goals of econometric analysis:
  - Estimating relationships between economic variables
  - Testing economic theories and hypotheses
  - Evaluating and implementing government and business policy
- Involves use of ***non-experimental & experimental*** data
- Also requires *appropriate* choice of econometric method(s)
- This, in part, will depend on the nature of the data utilised:
  - Cross-section data & pooled cross-section data
  - Time series data
  - Panel/longitudinal data

Data Handling & Cleaning

- Understanding the design and structure of data necessary for data cleaning and transformations of data.
- Data cleaning involves close scrutiny of data to detect and/or remove errors, inconsistencies and duplication of records
  - Requires detailed data analysis
  - Ensures data integrity and quality of data
  - Knowledge gained here informs economic & econometric modelling

Data Transformations

- Common to generate transformations of variables when working with economic data
  - Transform levels to growth rates
    - e.g. Change in GDP between periods
  - Construct new variables to capture economic outcomes
    - e.g. GDP per capita used to compare incomes of different countries
  - Use indices to summarise or adjust economic data
    - Convert information regarding many items into one index
      - e.g. FTSE 100, Dow Jones
    - Deflate economic series (convert nominal values to real values)
      - e.g. incomes, wages, output

Data Transformations

- Common to adjust economic data for effects of inflation
- Economics concerned with changes in REAL variables – requires adjustment for inflationary processes
- Important to identify relevant index with which to deflate your economic data
- Indices may be rebased or replaced to commence at a new point in time
  - May require you to splice indices to obtain consistent series over longer period of time
- Important to be consistent when deflating data
  - Choosing a common base period for all your data will provide for more informative and meaningful descriptive analyses
  - May also wish to consider regional (spatial) dimension

Data Transformations

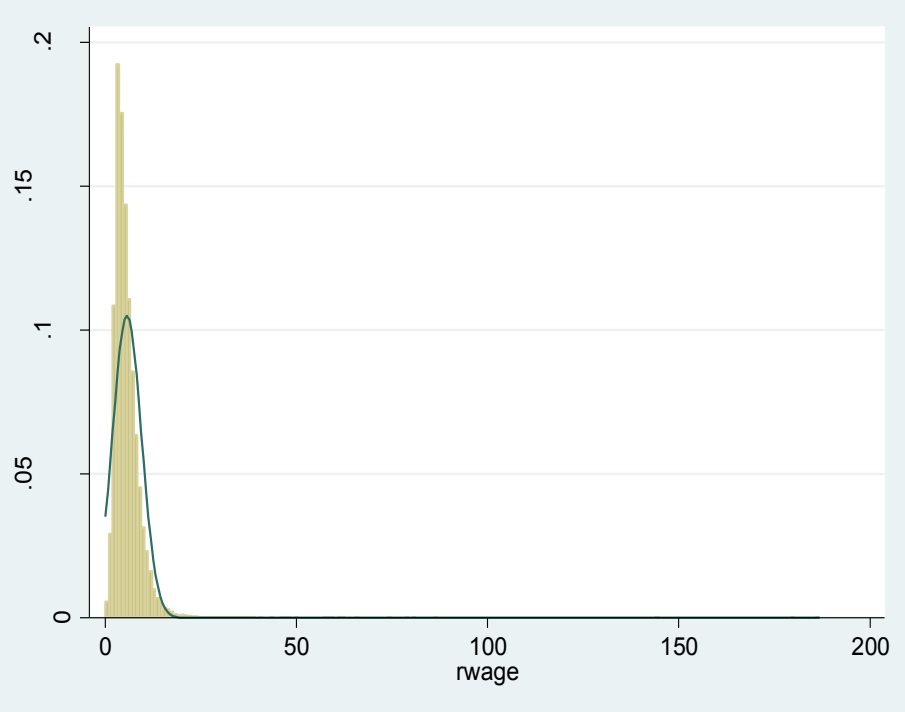
- Common transformation in economics: natural logarithm (where base e = 2.71828)
- Useful where data exhibit constant growth rate
- Frequently applied in both time-series and cross-section economics research
  - May facilitate greater symmetry of data and make shape more Gaussian (normal)
  - Facilitates identification of outliers, particularly for skewed data
  - Facilitates fitting of linear models – multiplicative processes become additive
  - Useful for re-expressing the scale of measurement and for approximating growth rates (changes in levels)

Data Transformations

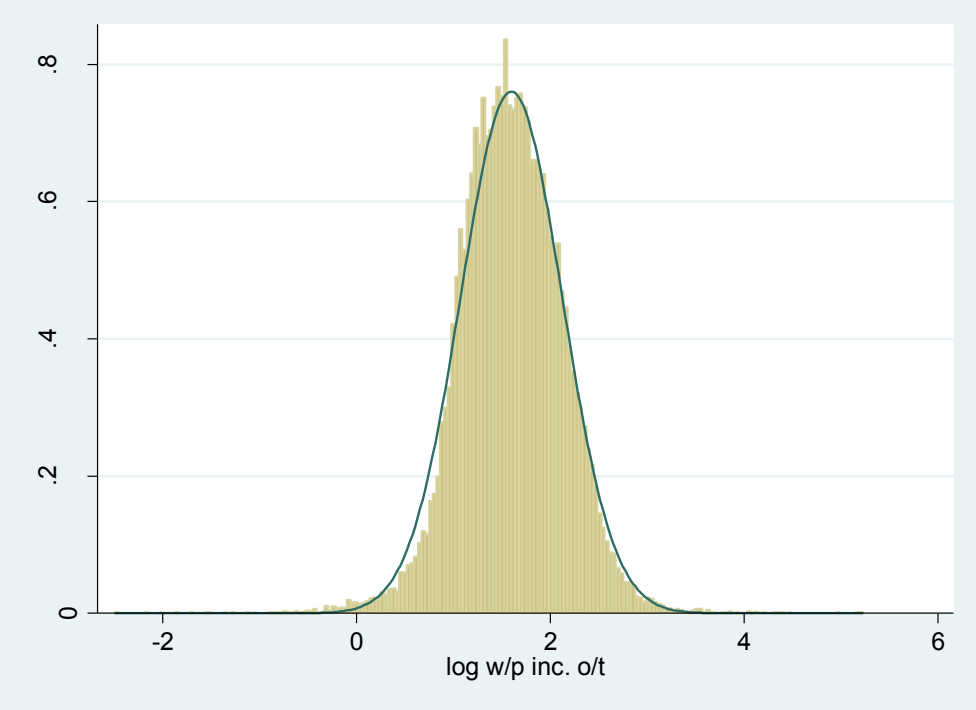
- Gaussian distribution cornerstone of many statistical tests and applications
  - Symmetric bell-shaped distribution with fixed proportions of the distribution at different distances from the centre
- Useful distribution in that it can be defined by its mean and standard deviation
  - Can reconstruct exact shape of the curve using this information
  - Can calculate the proportion of the area under the curve falling between various points
  - 95% of data lie within 2 SD of mean for standardised Gaussian
- **Many empirical distributions described by Gaussian shape but many are not. Nonetheless, serves as a benchmark comparator**
  - Can better compare distributions with different shapes once they have been transformed to approximate the Gaussian distribution

Data Transformations

Real Wage



Log Real Wage



Data Transformations

- Natural logarithms useful for constructing growth rates
- The **Annual Compound Growth Rate** (AGR) takes into account that the base for growth continues to rise over time
  - This is analogous to the calculation of compound interest
  - On this basis: $Y_t = (1 + g)^x \cdot Y_0$
- where Y_t = current Y, g = growth rate, x = no. of periods, Y_0 = initial Y
  - Rearranging yields the AGR:

$$g = \left(\frac{Y_t}{Y_0}\right)^{\frac{1}{x}} - 1$$

Data Transformations

Can approximate AGR by using **natural logs**

- Recall that $Y_t = (1+g)^x \cdot Y_0$
- Taking natural logs of the above equation we obtain:

$$\ln(Y_t) = x \cdot \ln(1+g) + \ln(Y_0)$$

- which rearranges to

$$\ln(1+g) = \frac{\ln(Y_t) - \ln(Y_0)}{x}$$

- Using our approximation:

$$g \approx \frac{\ln(Y_t) - \ln(Y_0)}{t}$$

- This yields the instantaneous growth rate gY

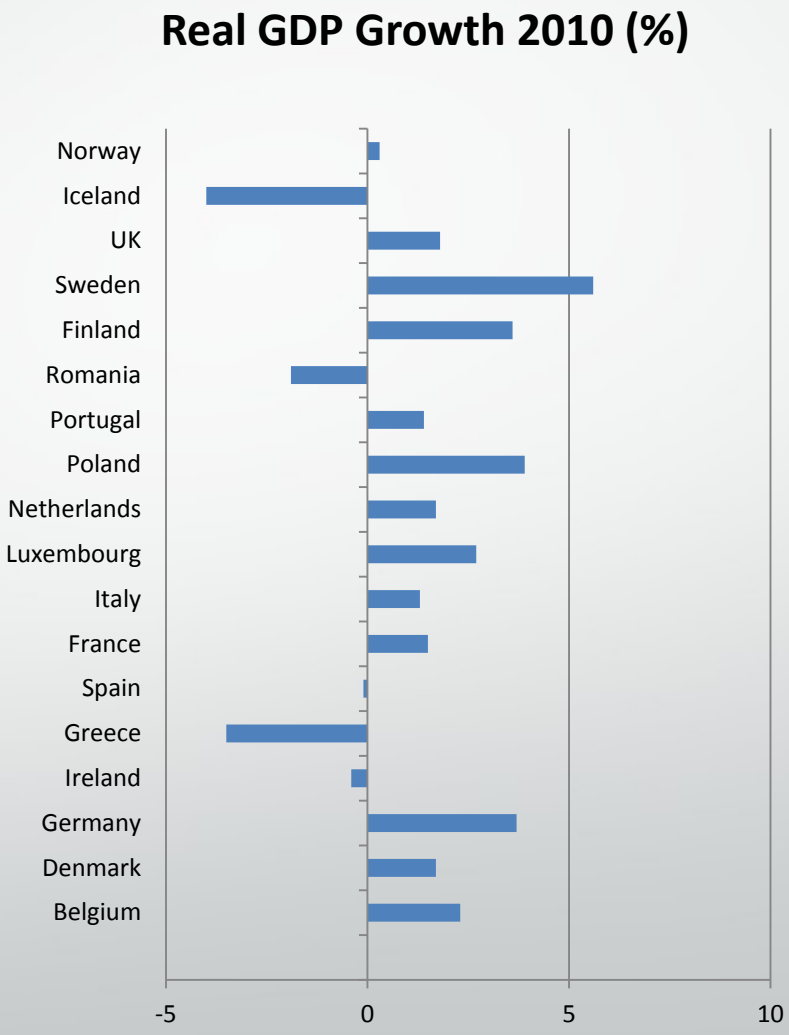
Describing Data

- Important to describe data and to present information in a clear, concise and accurate manner
- Variety of methods can be used to provide a descriptive analysis of data:
- Range of techniques summarised by:
  - Graphical Analyses
  - Numerical Analyses
- Key design features for these methods is:
  1. They tell us something about the underlying data
  2. They are reasonably familiar to people and easy to understand
- Important that the such methods do not distort the underlying evidence contained in the data

Bar Chart

- Frequently used for displaying observations over time or under different conditions
- Used to plot discrete (or categorical) data which has discrete values
- Requires small data sets that can be summarised easily
- The bars can be plotted vertically or horizontally
- Bars can also be stacked or set side-by-side
- Look similar to histograms but should not be mistaken for them!

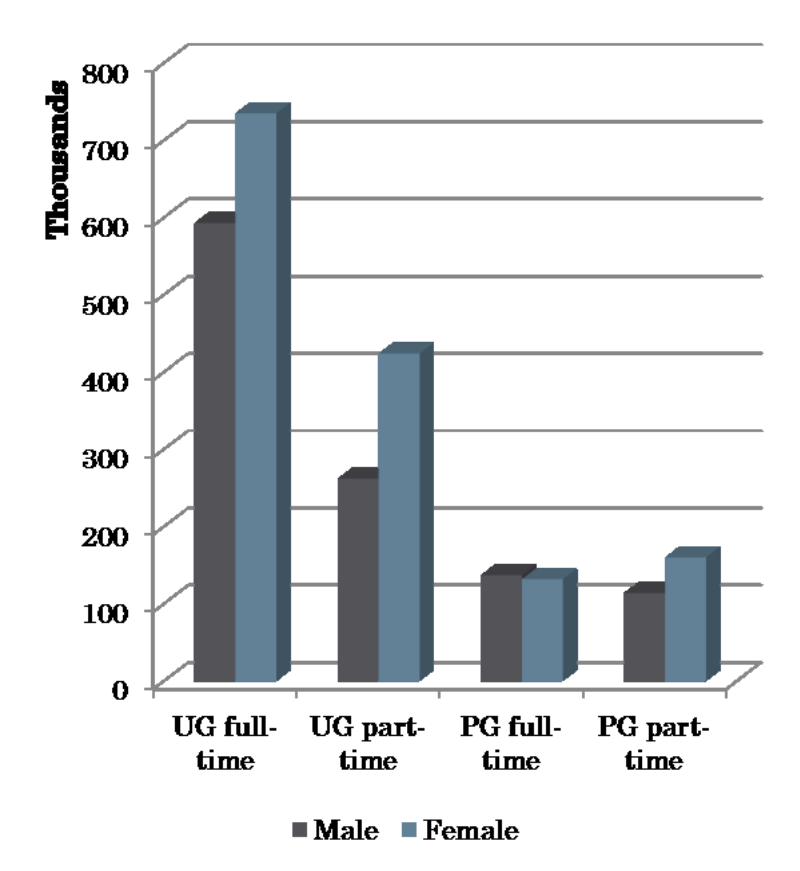
Bar Chart (horizontal)



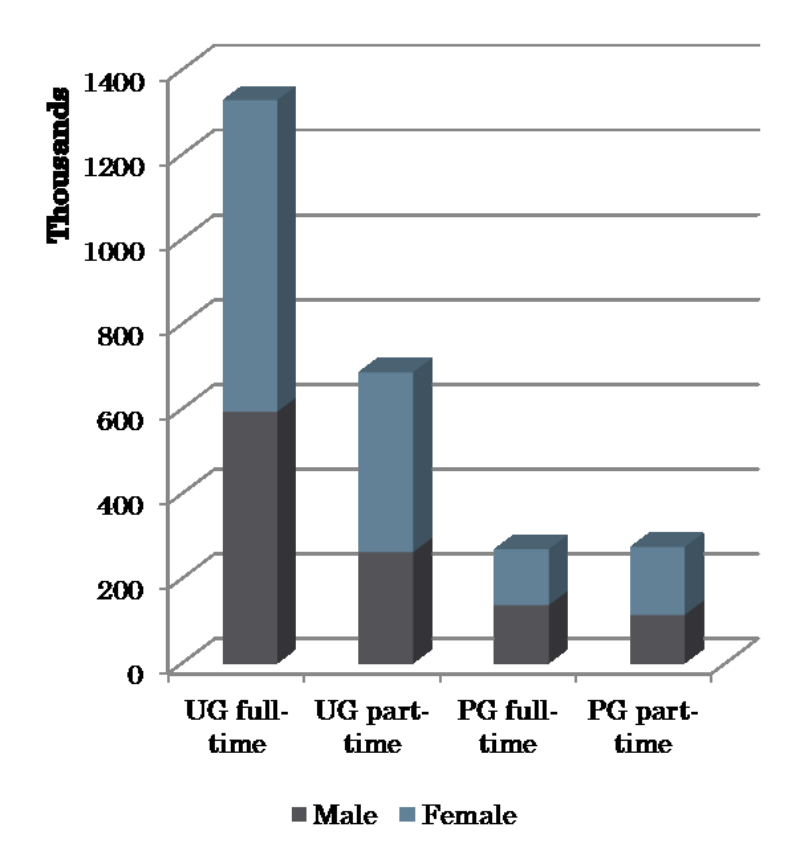
Source: OECD

Bar Chart (vertical & stacked)

Students in HE 2009 (vertical)



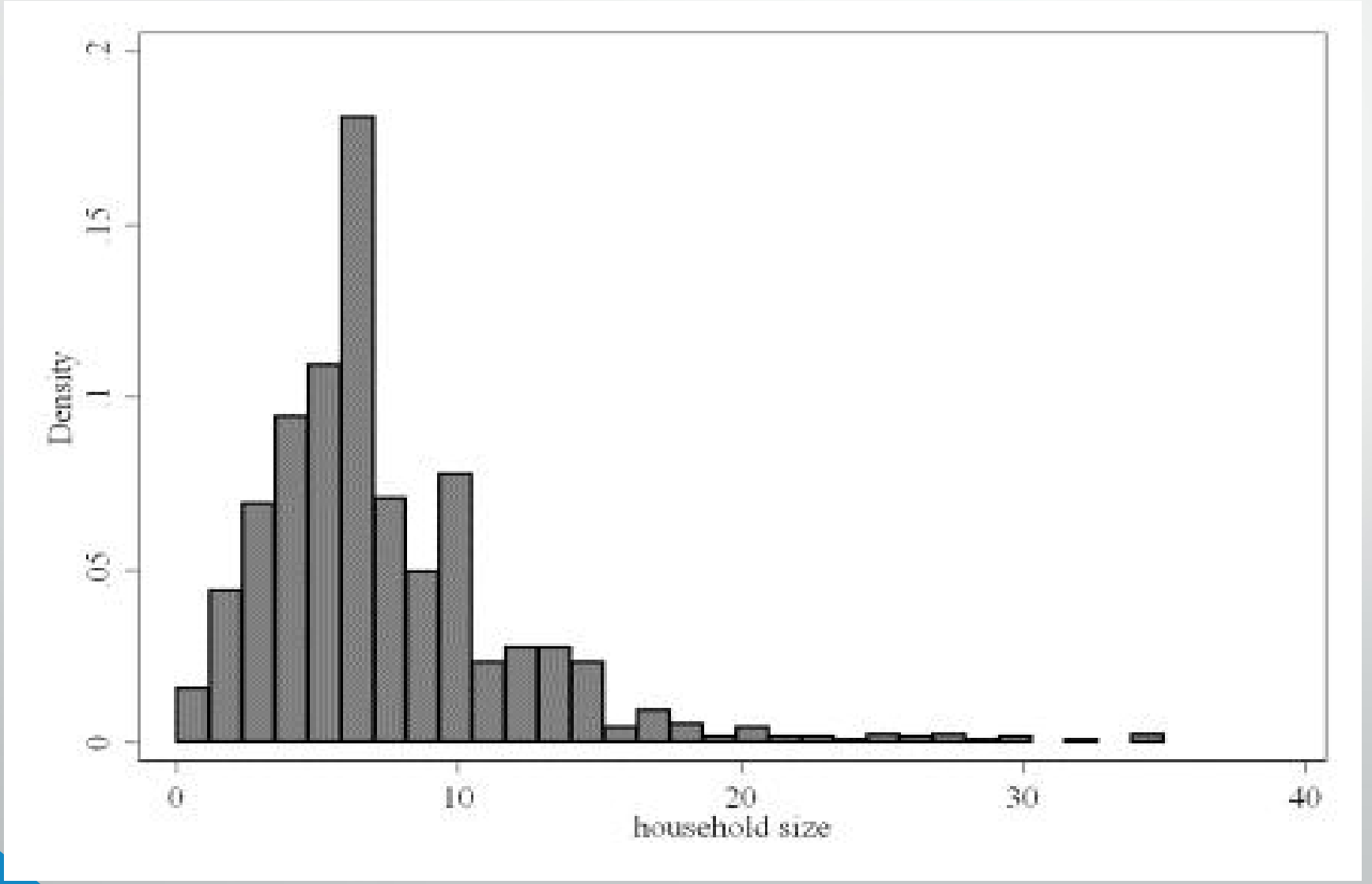
Students in HE 2009 (stacked)



Histogram

- A histogram is similar to a bar chart except that it corrects for differences in class interval sizes
- Where class intervals differ, bar charts give a misleading impression of the frequency distribution of the data
- A histogram plots frequencies against class intervals
  - Achieves this this by making the area of each bar represent its class frequency
  - Hence, for a given class frequency, if the class interval is twice as wide, then the bar will be half as tall
- Histograms provide important information about the shape of a distribution

Histogram

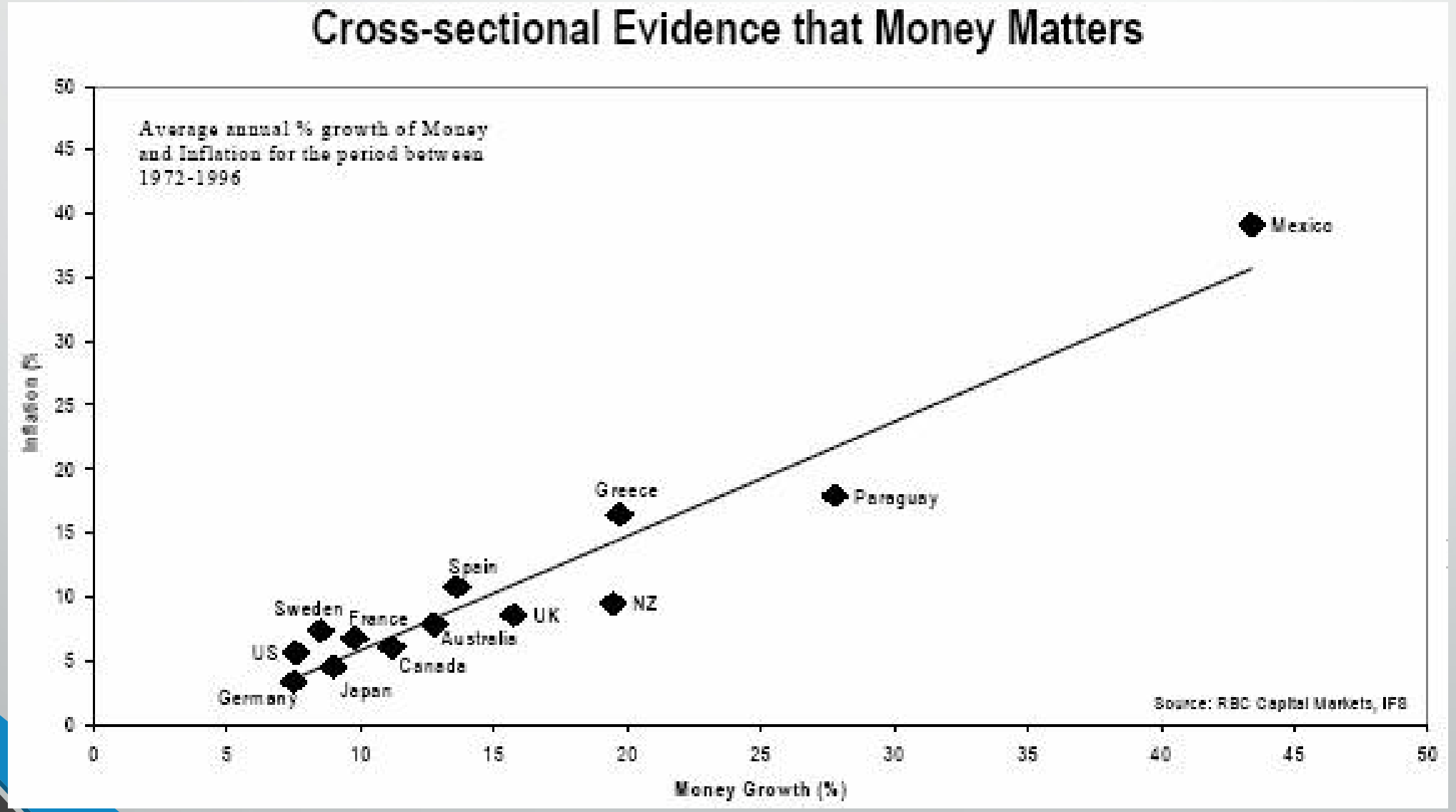


XY Scatter

- Often interested in the nature of relationships between two or more variables
  - e.g. money growth and inflation, education and employment
- XY scatter diagrams useful in this regard
- Provide **a quick visual impression of the relationship**
  - May observe a positive relationship where high values of one variable are associated with high values or another variable
  - May observe a negative relationship where high values of one variable are associated with low values or another variable
  - May observe no relationship between the two sets of values
- Important to note that there will often be exceptions to observed tendencies of the data

  Also, such relationships may or may not be causal

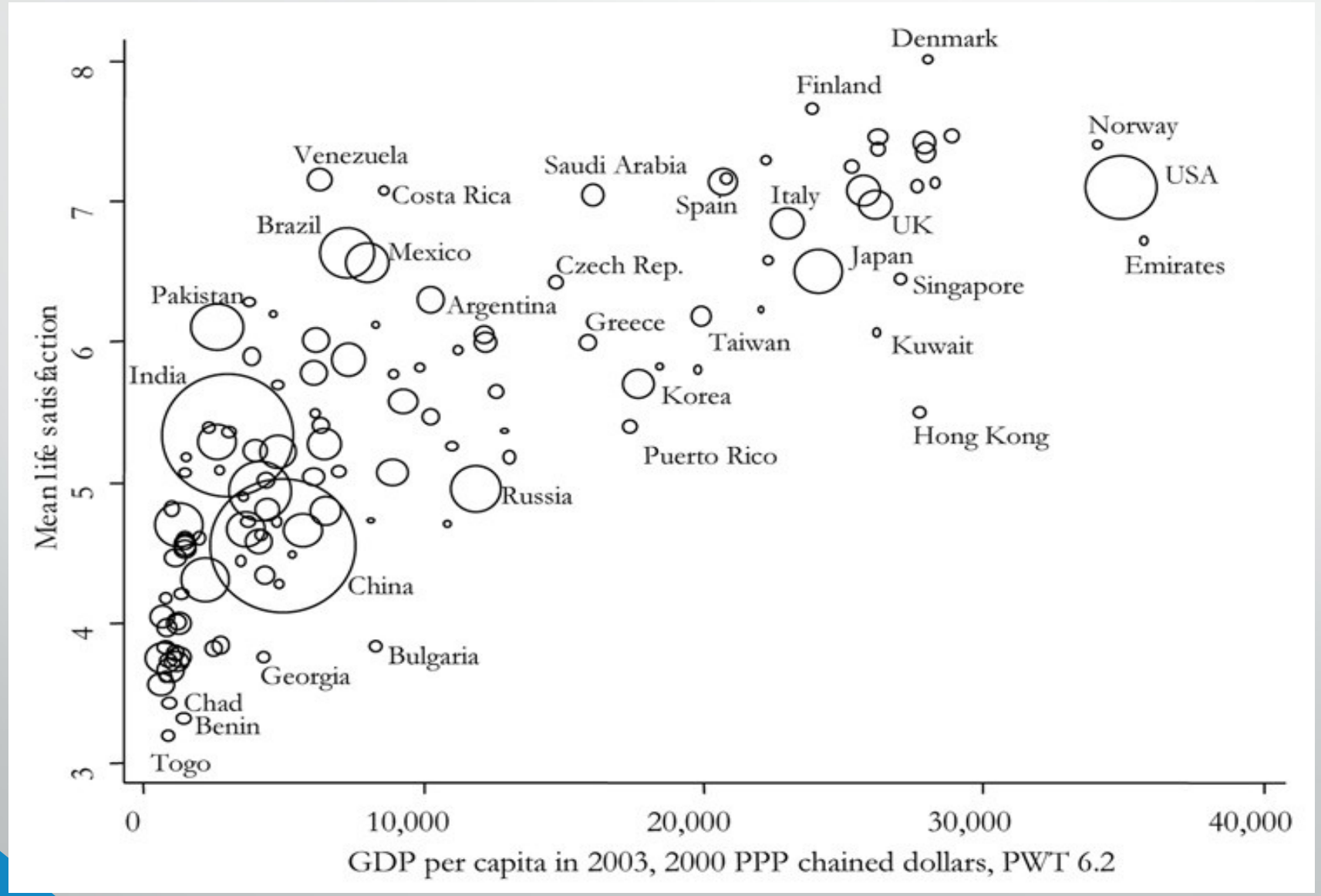
XY Scatter


Cross-sectional Evidence that Money Matters
Average annual % growth of Money and Inflation for the period between 1972-1996
Inflation (%)
Money Growth (%)
Mexico
Paraguay
Greece
Spain
NZ
UK
Sweden
France
Australia
US
Canada
Japan
Germany
Source: RBC Capital Markets, IFS
0
5
10
15
20
25
30
35
40
45
50

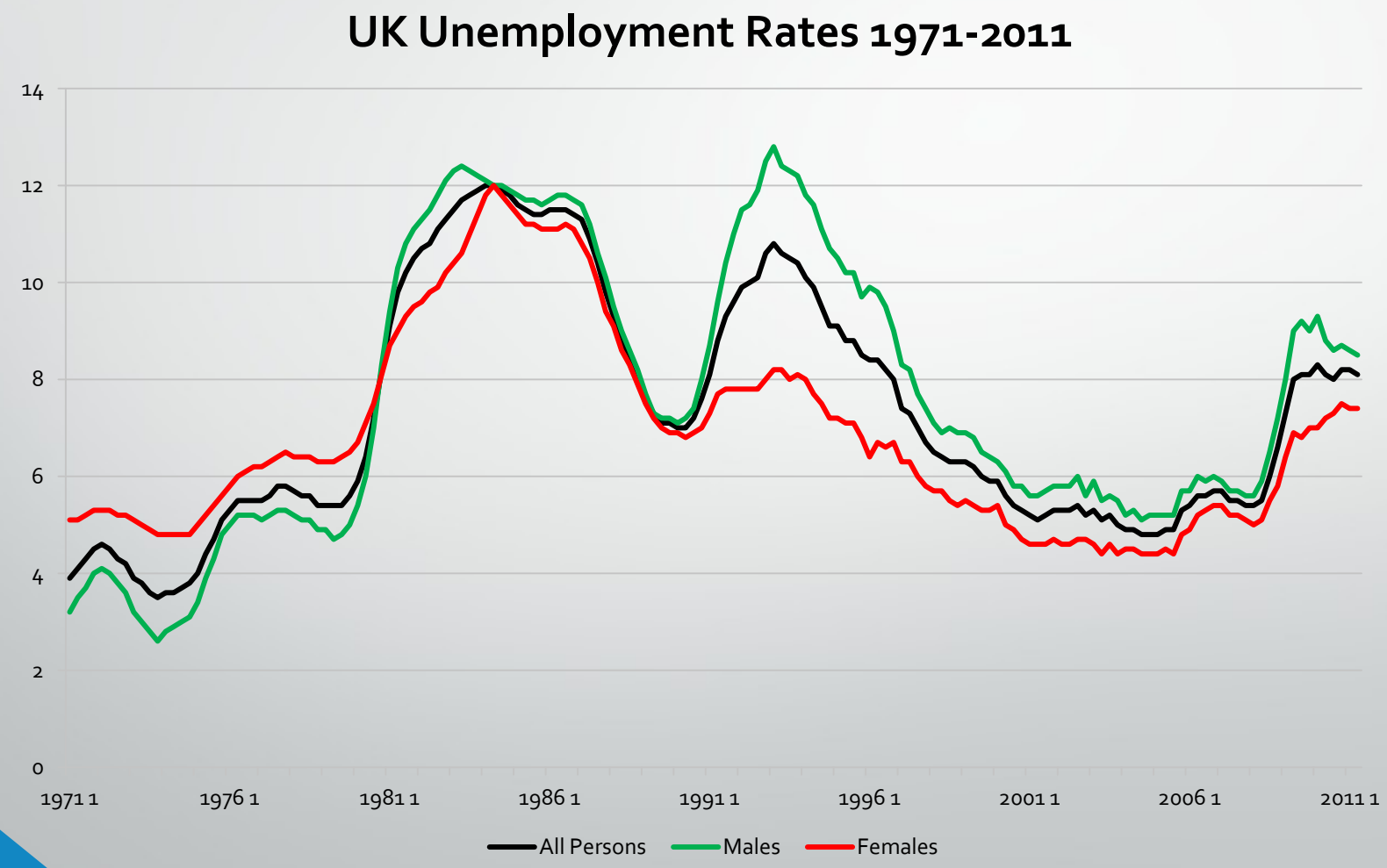
Bubble Chart (3D in 2D)

- A Bubble chart is a variation of a Scatter graph in which the data points are replaced with bubbles
- Commonly used when data has several series each of which contains a set of values you wish to illustrate
- Useful when you wish to compare series in terms of both their size and their relative position
  - Both X and Y axis of the bubble chart are numeric scales such that the position of plot is an indicator of two distinct numeric values
  - The **area of the plot** depends on the magnitude of a third set of numeric values
- Bubble charts are often used to display a wide range of financial and macroeconomic data

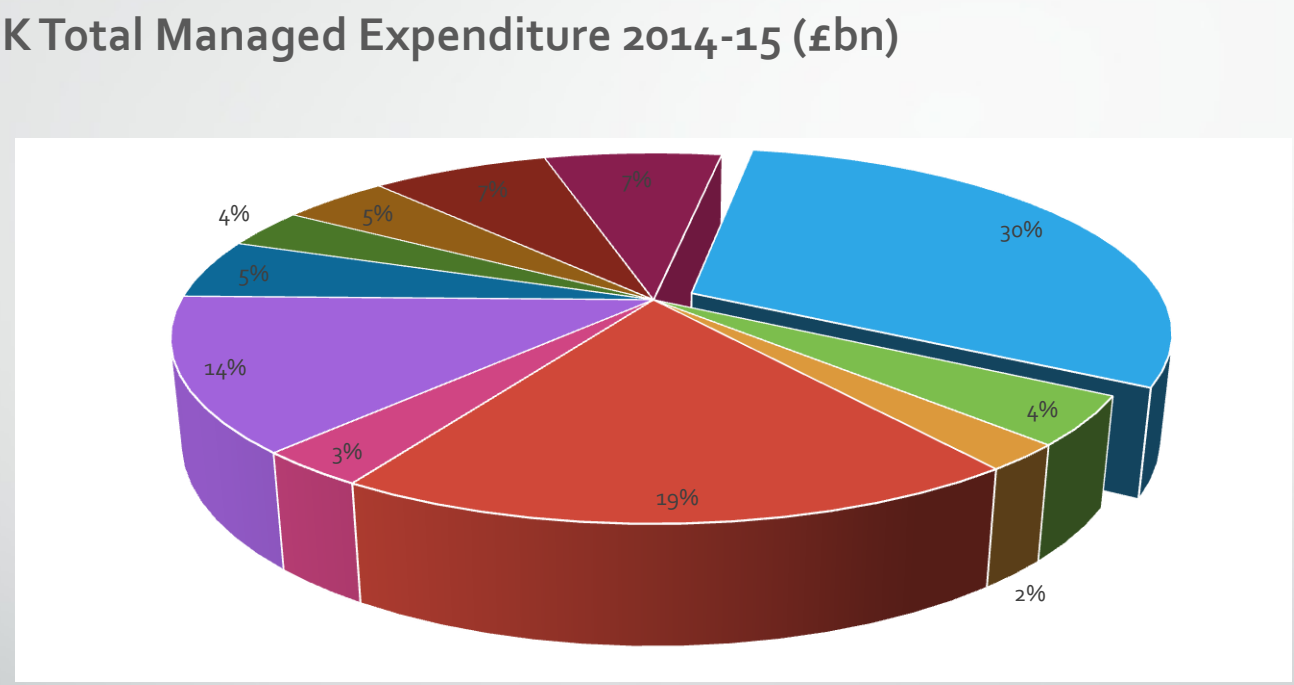
Bubble Chart



Line Graph


UK Unemployment Rates 1971-2011
14
12
10
8
6
4
2
0
1971 1
1976 1
1981 1
1986 1
1991 1
1996 1
2001 1
2006 1
2011 1
All Persons
Males
Females

Pie Chart


UK Total Managed Expenditure 2014-15 (£bn)
30%
4%
2%
19%
3%
14%
5%
4%
5%
7%
7%
Social protection
Personal social services
Industry, agriculture and employment
Health
Transport
Education
Defence
Housing and environment
Public order and safety
Debt interest
Other