# Week 4

# http://www.bit.ly/soci832_2019_week1

**Milena**

```
European_union<- as_tibble(European_union)

view_df(European_union, show.frq = TRUE, show.na = TRUE, max.len = 300)

European_union <- rec(European_union, v66, rec = "1=5; 2=3; 3=1;4=3; NA=3",
            append = TRUE)
```

**Jack**

```
#recode SUICIDE2
kgss2009.1 = rec(kgss2009.1, SUICIDE2, rec = "1=5; 2=4; 3=3; 4=2; 5=1; 8=NA; NA=NA",
append = TRUE)
kgss2009.1$SUICIDE2_r = set_labels(kgss2009.1$SUICIDE2_r, labels = c(
  "Strongly disagree" = 1,
  "Disagree" = 2,
  "Neither agree nor disagree" = 3,
  "Agree" = 4,
  "Strongly agree" = 5))
#test recode SUICIDE2
frq(kgss2009.1$SUICIDE2_r)
frq(kgss2009.1$SUICIDE2)
xtabs(~SUICIDE2_r+SUICIDE2, data=kgss2009.1)

#Widowed dummy
kgss2009.1 = rec(kgss2009.1, MARITAL, rec = "2=1; 1=0; 3:8=0", append = TRUE)
kgss2009.1$MARITAL_r2 = set_label(kgss2009.1$MARITAL_r2, label = "Widowed")
kgss2009.1$MARITAL_r2 = set_labels(kgss2009.1$MARITAL_r2, labels = c(
                                                    "Widowed" = 1,
                                                   "Not Widowed" = 0))
kgss2009.1$MARITAL_r2 = to_factor(kgss2009.1$MARITAL_r2)
#test Widowed dummy
frq(kgss2009.1$MARITAL_r2)
xtabs(~MARITAL_r2+MARITAL, data=kgss2009.1)

#recode SATFACE6
```

```
frq(kgss2009.1$SATFACE6)
kgss2009.1$SATFACE6_r = 6 - kgss2009.1$SATFACE6
kgss2009.1$SATFACE6_r = set_labels(kgss2009.1$SATFACE6_r, labels = c(
  "Very satisfied" = 5,
  "Very dissatisfied" = 1))
#test recode SATFACE6
frq(kgss2009.1$SATFACE6_r)
frq(kgss2009.1$SATFACE6)
xtabs(~SATFACE6_r+SATFACE6, data=kgss2009.1)

#create univariant table
print(descr(as_numeric(kgss2009.2),
            omit.headings = TRUE,
            stats = c("min", "max", "mean", "sd"),
            transpose = TRUE),
      method = "browser",
      footnote = NA)

#create bivariate matrix
sjp.corr(as_numeric(corr.kgss2009.2))
```

**Emma**

```
ds <- rec(ds, h1, rec = "1 = 1; 2 = 0",
          append = TRUE)

ds$h1_r <- set_label(ds$h1_r, label = "Gender")
ds$h1_r <- set_labels(ds$h1_r, labels = c ("Male" = 1,
                                           "Female" = 0
                                           ))
frq(ds$h1_r)

Gender (x) <numeric>
# total N=3955  valid N=3878  mean=0.48  sd=0.50

 val  label  frq raw.prc valid.prc cum.prc
   0 Female 2023   51.15      52.17   52.17
   1   Male 1855   46.90      47.83  100.00
  NA     NA   77    1.95         NA      NA



ds <- rec(ds, h12attch, rec = "1 = 1; NA = 0",
          append = TRUE)
```

```
ds$h12attch_r <- set_label(ds$h12attch_r, label = "Sent email attachment")
ds$h12attch_r <- set_labels(ds$h12attch_r, labels = c ("Yes" = 1,
                                    "No" = 0
))

frq(ds$h12attch_r)

Sent email attachment (x) <numeric>
# total N=3955  valid N=3955  mean=0.74  sd=0.44

 val label  frq raw.prc valid.prc cum.prc
   0    No 1019   25.76    25.76   25.76
   1   Yes 2936   74.24    74.24  100.00
  NA    NA    0    0.00       NA      NA



ds <- rec(ds, h12mlmed, rec = "1 = 1; NA = 0",
          append = TRUE)
ds$h12mlmed_r <- set_label(ds$h12mlmed_r, label = "Posted a,v,i files")
ds$h12mlmed_r <- set_labels(ds$h12mlmed_r, labels = c ("Yes" = 1,
                                        "No" = 0)
frq(ds$h12mlmed_r)
```

**Paul**

```
frq(auto$q1_a)

Q1_A. Public transport in [CITY NAME], for example the bus, tram or metro - Gene
(x) <numeric>
# total N=37626  valid N=37626  mean=2.83  sd=2.32

 val              label   frq raw.prc valid.prc cum.prc
   1       Very satisfied  8557   22.74     22.74   22.74
   2     Rather satisfied 17074   45.38     45.38   68.12
   3   Rather unsatisfied  4925   13.09     13.09   81.21
   4 Not at all satisfied  2924    7.77      7.77   88.98
   9               DK/NA  4146   11.02     11.02  100.00
  NA                 NA     0    0.00       NA      NA
```

Richard L

```
> nsw_crime <-rec(nsw_crime, hsdegree, rec ="0:50 = 1; 51:70 = 2; 71-100 = 3",
append = TRUE)
```

```
> nsw_crime$hsdegree_r <- set_label(nsw_crime$hsdegree_r, label = "Secondary
Education")
> nsw_crime$hsdegree_r <- set_label(nsw_crime$hsdegree_r, label = "Secondary
Education")
> nsw_crime$hsdegree_r <- set_labels(nsw_crime$hsdegree_r, labels = c ("Low" =
1, "Medium" = 2, "High" = 3))

frq(nsw_crime$hsdegree_r)

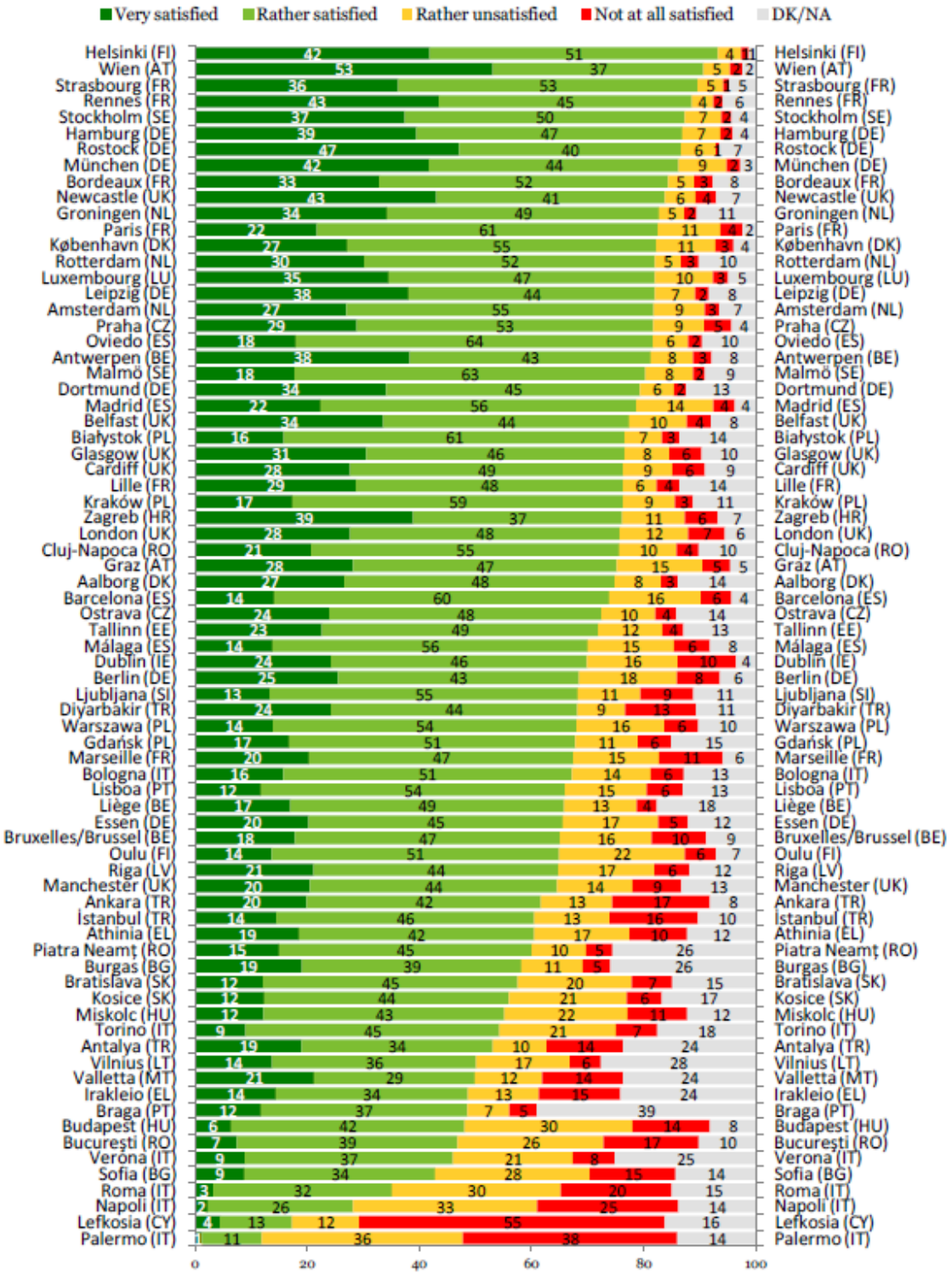
Secondary Education (x) <numeric>
# total N=129  valid N=120  mean=1.15  sd=0.36

 val  label frq raw.prc valid.prc cum.prc
   1    Low 102   79.07        85      85
   2 Medium  18   13.95        15     100
   3   High   0    0.00         0     100
  NA     NA   9    6.98        NA      NA

> descr(nsw_crime$hsdegree)
Descriptive Statistics
nsw_crime$hsdegree
Label: % of residents who completed year 12 or equivalent
N: 129

                  hsdegree
----------------- ----------
             Mean      38.46
          Std.Dev      14.61
              Min      21.20
               Q1      28.00
           Median      31.80
               Q3      45.00
              Max      76.00
              MAD       7.26
              IQR      16.45
               CV       0.38
         Skewness       1.18
      SE.Skewness       0.21
         Kurtosis       0.16
          N.Valid     127.00
        Pct.Valid      98.45
```

## Satisfaction with public transport




**Q1. Generally speaking, please tell me if you are very satisfied, rather satisfied, rather unsatisfied or not at all satisfied with each of the following issues:**

Base: all respondents, % by city

# Week 3

# http://www.bit.ly/soci832_2019_week1

Name, Article, Concepts.

Instructions

1. Identify examples of the following concepts in one of the example articles below.
2. Paste examples with brief explanations of how the examples illustrate the concepts you find into the Google Doc here.

Emma, McAllister (2016), population parameter, null hypothesis, p-value, effect size,

- Population parameter: “The data come from the 2013 AES, which was a nationally representative mail-out, mail-back survey of persons registered to vote in the 2013 election. The sampling frame was supplied by the Australian Electoral Commission from the electoral rolls. There was also an online option for completion of the survey, which was used by a small number of respondents. The final response rate was 33.9% after four follow-ups. The survey was weighted to reflect the characteristics of the national electorate.”
- Null hypothesis: Article doesn’t explicitly specify null hypotheses, but does include four hypotheses which are either supported or rejected (which we can assume was done by rejecting a null hypothesis)
- Effect size e.g. Adj R2

Jack; Harrigan 2017; Dimension reduction and finding categories: Factor analysis, Cluster analysis

Foreign ownership - conservatives/labor/bipartisan donation
company has a director on a government board => making a bipartisan donation
Being anti-regulation => donating conservatives
(p. 740)

# MILENA: Internet use, political knowledge and youth electoral participation in Australia

Regresion model,  SE and P values Pag. 11 and 12
18- 24 years old = sample
All australian= population

Li, McAllister, sample statistics.
Any measurement (eg. mean, regression coefficient estimates) that describes the sample of data that is used to infer about the whole population.
[Table 2 summarises sample statistics]

Richard Lacinski
Harrigan (2017), Descriptive Statistics for Total Corporate Political Donations
Mean: (A$) Was $5896
Standard Deviation: $31178 AUD
Min/Max: 0/495127
N = 1575 (143 were non-zero observations)

Table 8:  Models of donation behaviour (regulatory environments theory variables and simple controls)
Confidence Interval:
P-value :
Effect Size:
Standard Error:

**Akbar, Harrigan (2017), Concepts:**

**Descriptive Statistics**
Mean: For example, the mean donation of the average Top 2000 corporation who made a donation (AU$40,000) was less than the average male salary and only 0.005 per cent (or 1/20,000th) of the average corporations' annual turnover.

Pearson correlation (r ) for outcomes variables = page 739

R Square = page 743 R-square on the model of donation behaviour

SD: Table III

**Inferential Statistics**

OLS regression with regression logistic: page 738

18-24 year olds from sample
Political knowledge for age groups. Sample mean = sample statistic. Figure shows 95 % confidence interval: IIkelihood that population mean is represented within is 95%
Used to create tables 2 and 3, showing probability of possessing political knowledge for table 2.
For variables listed, likelihood of having political knowledge. Politically interested (on scale 1 to 4) more likely to have political knowledge between scale points (0.857), + or - 0.138
Controlling for political interest etc. election internet use frequency impacts political knowledge (0.418)
*Statistically significant at p<.01 so 99 times out of 100
TL

Jarvis, Harrigan (2017), Table VII, Number of obs. (n) 1575, it means the sample statistic used in this regression model is 1575.

Masa, Harrigan (2017), population of data: The primary unit of analysis is a subset of the IBISWorld Top 2000 Companies: the 1,575 for profit organization.

H1 = using the internet to increase political knowledge among the young can result in higher levels of electoral participation

H0 =  no relationship between using the internet to increase political knowledge among the young and higher levels of electoral participation

# Week 2

# http://www.bit.ly/soci832_2019_week1

Exercise 2

Can you identify the following concepts in our model/demonstration article:

- theory & hypotheses  Milena
- units of analysis and variable . Milena
- dependent variable, independent variables, control variables Paul
- conceptualisation of a variable Masa
- operationalisation of a variable Jack Woo
- one of each type of variable: categorical, binary, ordinal, continuous (or interval) Jarvis
- a scale or index -- Akbar
- the source for the scale or index Li
- evidence the scale, index, or other variables are reliable and valid Tian

Theory: the internet has the capacity to increase political knowledge and thereby to strengthen democratic citizenship

Units of analysis: survey respondents(3955)  and young people(404)

Hypothesis

> Hypothesis 1 (H1): Following an election on the internet will lead to higher levels of political knowledge, other things (including other media sources) being equal.
> Hypothesis 2 (H2): The impact on political knowledge of following an election on the internet will be greater among the young compared to the rest of the electorate, other things being equal.

**Hypothesis 3 (H3): Higher levels of political knowledge will result in higher levels of electoral participation, other things being equal.**
**Hypothesis 4 (H4): The impact of political knowledge on electoral participation will be greater among the young compared to the rest of the electorate, other things being equal.**

### Dependent variables

> Two dependent variables are used, political knowledge and the likelihood of voting. Political knowledge is measured by the proportion of correct answers to 10 factual questions about politics included in the 2013 AES survey. Six of the questions related to the operation of the political system in general, three to the circumstances of the 2013 election, and one to international politics. The 10 items were combined into an additive scale, producing a mean of 4.5 (see Appendix 1).

## Independent Variables

Three variables are used to measure internet use. Frequency of general internet use is measured by responses to the question: 'In general, how often do you use the internet?' and is coded from 1 (never uses the internet) to 7 (uses the internet several times per day). Frequency of election internet use is measured by responses to the question: 'Did you make use of the internet at all to get news or information about the 2013 federal election?' and is coded from 1 (never used) to 5 (used many times). Internet skills is an additive scale based on the question: 'Have you done any of the following tasks on the internet? … sent an attachment via email … downloaded a software program. … posted audio, video or image files. … personally designed a webpage or blog.' All three measures have been shown to tap into different aspects of how individuals used the internet, with consequences for political attitudes and behavior (van Deursen and van Dijk 2010; Gibson and McAllister 2014; Gil de Zúñiga, Jung, and Valenzuela 2012).

The variables measuring media sources were derived as follows. Frequency of following the election in the newspapers is based on the question: 'How much attention did you pay to reports about the election campaign in the newspapers – a good deal, some, not much or none at all?' Frequency of following the election on television and on the radio were based on the question: 'Did you follow the election campaign news on television – a good deal, some, not much or none at all? And did you follow the election campaign news on the radio?' All three variables are coded from 1 (none at all) to 4 (a good deal). These questions provide a summary measure of self-reported media use during the election campaign.

Six socioeconomic variables are used as controls in the models. Gender, tertiary education and Australian born are dummy variables. Gender is used as a control variable since there is evidence that women and men have different rates of internet usage and often use the internet in different ways (see, e.g. Helsper 2010). Education discriminates in the quality and quantity of political information that a voter seeks and is therefore also an important control variable. Birthplace is taken into account since one in five of the Australian population have been born overseas, about half of them in non-English speaking countries, and their political outlooks often differ from the Australian born (Bilodeau, McAllister, and Kanji 2010). Age, which reflects cumulative political life experiences, is measured in years and family income is measured in quintiles. Urban residence is a four point scale coded from 1 (lives in a rural area or village) to 4 (lives in a major city over 100,000 people). This is a potential important control variable since the speed and availability of internet access varies considerably across Australia.

One attitudinal variable, interest in politics, is also included as a control. It is measured by the question: 'Generally speaking, how much interest do you usually have in what's going on in politics?' and is coded from 1 (none) to 4 (a good deal). Political interest is an important control since it is one of the most significant predictors in its own right of political knowledge (Gronlund and Milner 2006), as well as of electoral participation (Blais and Rubenson 2013). It has also been shown to be stable over time, and to have its roots in childhood political socialization and adolescence (Prior 2010). Given the importance of political interest, it is therefore crucial to take it into account to properly specify the predictive model.

Jarvis:
Categorical:
Gender.

Binary:
Internet skills is an additive scale based on the question: ‘Have you done any of the following tasks on the internet? … sent an attachment via email … downloaded a software program. … posted audio, video or image files. … personally designed a webpage or blog.’

Ordinal:
Since voting in Australia is compulsory, the second dependent variable, electoral participation, is based on an attitudinal rather than a behavioral question: ‘Would you have voted in the election if voting had not been compulsory?’ and is scored from 1 (definitely would not have voted) to 5 (definitely would have voted).

Continuous:
Age

**Akbar - Scale or Index?**

| **Variable** | |
|---|---|
| Political Knowledge | Index - the variable is summed up from some indicators |
| The Likelihood of Voting | Scale:<br>“Would you have<br>voted in the election if voting had not been compulsory?” |
| Internet Use | Index<br><br>Measured by three indicators -> Frequency of general internet use, Frequency of election internet use, Internet skills |

**Source of scale (Li)** - data source is the AES

**Appendix 1: Variables, scoring and means.**

| Variable | Coding | Means | |
|---|---|---|---|
| | | All | 18–24 |
| *Likelihood of voting* | From 1 (definitely not vote) to 5 (definitely vote) | 4.20 | 3.70 |
| *Political knowledge* | From 0 to 10 | 4.51 | 3.41 |
| *Internet use* | | | |
| Frequency of general internet use | From 1 (never uses) to 7 (several times per day) | 5.77 | 6.75 |
| Frequency of election internet use | From 1 (never used) to 5 (used many times) | 2.66 | 3.07 |
| Internet skills | Additive scale, 0–4 | 2.14 | 3.23 |
| *Media sources* | | | |
| Frequency of following election in newspapers | From 1 (none at all) to 4 (a good deal) | 2.47 | 2.10 |
| Frequency of following election on TV | From 1 (none at all) to 4 (a good deal) | 2.90 | 2.63 |
| Frequency of following election on radio | From 1 (none at all) to 4 (a good deal) | 2.31 | 2.00 |
| *Controls* | | | |
| Political interest | From 1 (none) to 4 (a good deal) | 3.02 | 2.56 |

Richard Lacinski

Theory: McAllister's theory is that internet use shapes the political knowledge of young people affecting their overall political engagement

Hypotheses:
Hypothesis 1 (H1): Following an election on the internet will lead to higher levels of political knowledge, other things (including other media sources) being equal.

Hypothesis 2 (H2): The impact on political knowledge of following an election on the internet will be greater among the young compared to the rest of the electorate, other things being equal.

Hypothesis 3 (H3): Higher levels of political knowledge will result in higher levels of electoral participation, other things being equal.

Hypothesis 4 (H4): The impact of political knowledge on electoral participation will be greater among the young compared to the rest of the electorate, other things being equal.

Variables:
Dependent Variables:
Two dependent variables were used which were, political knowledge and the likelihood of voting.

Independent Variables:
Three independent variables were used

1. Frequency of general internet use. Measured by the responses to the question "In general, how often do you use the internet?"
2. Frequency of election internet use is measured by responses to the question: "Did you make use of the internet at all to get news or information about the 2013 federal election?"
3. Internet skills is an additive scale based on the question: "Have you done any of the following tasks on the internet? … sent an attachment via email … downloaded a software program. … posted audio, video or image files. … personally designed a webpage or blog."

Jack Woo
Political knowledge: quiz
Electoral participation: attitude towards compulsory election ("attitudinal rather than a behavioral question: 'Would you have voted in the election if voting had not been compulsory (p. 5)?'")

Internet use: frequency of general/electoral internet use (How often...?)

Internet skills: tasks done on the internet
Media sources: frequency of following election on different forms of media

Education: completion of tertiary education
Birthplace: born in Australia/overseas

Conceptualisation of variables
Dependant variables: political knowledge - information about events, institutions or personalities – and 'background' or 'structural' political knowledge, which enables citizens to interpret political affairs.
Independent variables: internet use- Frequency of general internet use.

Reliability: "To examine these questions, this paper examines the role of the internet in shaping political knowl- edge among the young in the context of a national election campaign. In turn, extending other studies which have examined this topic (but cf Bakker and de Vreese 2011), the paper tests the consequences of political knowledge for the likelihood that a person will turnout to vote in a national election. The data come from the 2013 Australian Election Study (AES) survey which asked an extensive range of questions about political knowl- edge, in addition to detailed questions about internet use and electoral participation"

> mentions other similar studies, suggestion of repeatability and therefore reliability (would have to check whether actually replicated the same method though)

"As virtually all other studies have demonstrated, the level of political knowledge across the general electorate is consistently low (Table 1)."

> statement suggests that political knowledge is low generally, would be good to test before and after electoral campaign, and to measure the internet usage before and after the electoral campaign, and to measure the political knowledge before and after the electoral campaign. Difficult to really conclude internet shapes political knowledge during election campaigns and in turn affects electoral turnout. Also can you control for whether or not voting is compulsory?

Validity: "As Mondak (2001) has argued, a 'don't know' response may reflect a lack on information, while an incorrect response is more likely to indicate misinformation. To test for this possibility, the results were analyzed to see if 'don't know' and incorrect responses differed from one another in any systematic way. The results did not reveal any significant differences[6] and for that reason, the two categories are combined, with respondents being scored one if they gave a correct answer and zero if they gave an incorrect answer or said that they didn't know.[7]"

**Exercise 1**

1. Your name
2. Your topic
3. What datasets or articles have you found?
4. What lessons have you learnt from your search?

Li Tan
Politics/education/inequality
Macroeconomic data, election outcomes in Australian States & Territories, HILDA, Census
Find researchers - eg. Andrew Leigh - who publish in relevant areas, whose websites may contain articles, datasets for download and replication files.

Paul Gaynor
Public transport organizational form and customer satisfaction
Article is Fiorio et al "*User satisfaction and the organization of local public transport: Evidence from European citie*s".

Data set is a Flash Eurobarometer survey "Perception survey on quality of life in European cities".
https://data.europa.eu/euodp/en/data/dataset/S764_277_A_B_C_D

Jack Woo
Religion, stress and suicide acceptability in South Korea
2009 Korean General Social Survey
Centred around interesting topics and stick to general survey and census

WEEK 1

## Exercise 2: The best course you studied

What was the best course you ever studied? What made it so great?

- Please answer on the Google Doc next to your name.

Paul Gaynor - best course was on Program Evaluation, because I was using it at work before the course even ended.

Li - Graeco-Roman Studies. 3-week onsite summer course held in ancient sites around Greece and Italy

Jack - ANTH816 about medical anthropology, which teaches many practical examples of anthropological studies

Jarvis: SOCI 811 comparative social policy. A fantastic framework is introduced in this unit for me to compare social policy around the world.

Masa: SOC830 doing the social survey. Some classes are boring reading too much but SOC 830 was practical and fun to write codes.

Tian: Philosophy of the Environment (tutor was great, gave really great feedback, 5h@t all over my arguments, but was really constructive, learned a lot...or Science Fiction (lecturer was great, kind of crazy super into recommending movies and inappropriate theories and advice...or critical thinking (interesting and useful)

Richard Lacinski: A course that was interesting was Soc 863 Social Care and Human Services because I got to do my final assignment on something in my interests related to when I lived in Serbia.

Milena: Good class was qualitative research. Excellent structure, teacher was engaging, clear, link practical to theory.

**Exercise 1**

Name (Preferred name: ) Something about yourself? What do you do for work? Outside? Why are you enrolled in this class?

Paul Gaynor - I work for Transport for NSW. Research work in transport is increasingly about digital data, so a good knowledge of quantitative methods is increasingly important.

Tian - I own a three legged doggo - I work at NMI in the Policy section - Unit description sounded interesting and useful

Li; economist cum mum; wants to learn R to expand skills set (have heard it's really good once you get over the initial shock)

Fang 'Jack' Wu; studying development studies, focusing on the human rights issues in the Greater China; not working at this moment; wish to learn about the whole process of quantitative research.

Name: Richard Lacinski - Why I am enrolled in this class? Last semester I did the Social Research surveys class and found it interesting and I have an interest in potentially working in social/market research and though this class would usefully expand on my previously learned knowledge.
Outside of Uni: I work in an Agricultural research firm

Name: Jie (preferred name: Jarvis). After SOC830, I found quantitative research is interested. I want to practice my data analysis skills in this unit.

Name: Milena
Work: Work for NGO- Mental Health.

Name: Masanori Ito (Masa) - I want to improve quantitative research skills.

Nicholas Harrigan (Nick): Lecturer in Sociology at Macquarie University.