The [Australian Election Study](https://australianelectionstudy.org/) is an impressive long term research project that has collected the attitudes and behaviours of a sample of individual voters after each Australian federal election since 1987. All the datasets and documentation are freely available. An individual survey of this sort is a vital complement to the necessarily aggregate results provided by the Australian Electoral Commission, and is essential for coming closer to understanding political and social change. Many countries have a comparable data source.

Many thanks to the original researchers:

McAllister, Ian; Makkai, Toni; Bean, Clive; Gibson, Rachel Kay, 2017, “Australian Election Study, 2016”, doi:10.4225/87/7OZCZA, ADA Dataverse, V2, UNF:6:TNnUHDn0ZNSlIM94TQphWw==; 1. Australian Election Study December 2016 Technical Report.pdf

In this blog post I just scratch the surface of the latest available data, from the time of the 2016 federal election. In later posts I’ll get into modelling it a bit deeper, and hopefully get a chance to look at some changes over time. Some of the code I use in this post for preparing the data might find its way into the ozfedelect R package, but not the data itself.

**Getting the data**

The data need to be downloaded by hand from the Australian Data Archive via [the Dataverse](https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.4225/87/7OZCZA) (an international open source research data repository) in order to agree to the terms and conditions of use. The following code requires the SPSS version of the data and the Excel version of the data dictionary to have been saved in an ./aes/ subfolder of the working directory.

The SPSS data has column names truncated at eight characters long, which means that some of the variables in the data don’t match those in the data dictionary. The code below makes the necessary manual adjustments to variable names from the dictionary to deal with this, and re-classes the categorical responses as factors with the correct level labels in the correct order. I do this by importing and joining to the data dictionary rather than extracting the attributes information from the SPSS import, which might have also worked; I prefer my manual approach so I can deal with some anomalies explicitly (for example, many of the values listed as 999 “skipped” in the data dictionary are NA in the SPSS data).

*Post continues after code extract*

#====================Preparation=============

#----------------------------packages---------------

library(tidyverse)

library(scales)

library(rio)

library(svglite)

library(frs)

library(readxl)

library(testthat)

library(survey)

library(vcd)

library(RColorBrewer)

#---------------------Import and tidy data-----------------

# Download by hand from https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.4225/87/7OZCZA

aes2016\_orig <- import("aes/2. Australian Election Study, 2016.sav")

aes2016\_code\_orig <- read\_excel("aes/1. Australian Election Study, 2016 - Codebook.xlsx",

sheet = "Data Dictionary")

aes2016\_code <- aes2016\_code\_orig %>%

rename(question\_label = Label...3,

answer\_label = Label...5) %>%

rename\_all(tolower) %>%

fill(variable, position, question\_label) %>%

mutate(var\_abb = substring(variable, 1, 8)) %>%

mutate(var\_abb = case\_when(

var\_abb == "H19\_STAT" ~ "H19STATE",

var\_abb == "H19\_PCOD" ~ "H19pcoRV",

var\_abb == "H20\_othe" ~ "H20\_oth",

var\_abb == "H25\_othe" ~ "H25\_oth",

var\_abb == "Final\_ST" ~ "finSTATE",

var\_abb == "Samp\_STA" ~ "SamSTATE",

var\_abb == "detailed" ~ "doutcome",

var\_abb == "Responde" ~ "responID",

var\_abb == "Start\_Ti" ~ "Sta\_time",

var\_abb == "Samp\_PCO" ~ "SampcoRV",

var\_abb == "Final\_PC" ~ "finpcoRV",

var\_abb == "Total\_Du" ~ "totaldur",

TRUE ~ var\_abb

)) %>%

rbind(tibble(

variable = "StateMap", position = NA, question\_label = "Unidentified state variable",

value = 1:8,

answer\_label = c("New South Wales", "Victoria", "Queensland", "South Australia", "Western Australia",

"Tasmania", "Northern Territory", "Australian Capital Territory"),

var\_abb = "StateMap"

))

aes2016\_questions <- distinct(aes2016\_code, var\_abb, position, question\_label)

aes2016\_answers <- aes2016\_code %>%

select(var\_abb, variable, value, answer\_label) %>%

filter(!is.na(value))

# Check all the names in the data now match those in the data dictionary:

expect\_equal(

names(aes2016\_orig)[!names(aes2016\_orig) %in% aes2016\_questions$var\_abb],

character(0)

)

# ... and vice versa:

expect\_equal(

unique(aes2016\_questions$var\_abb)[!unique(aes2016\_questions$var\_abb) %in% names(aes2016\_orig)],

character(0)

)

aes2016 <- aes2016\_orig %>%

as\_tibble()

attributes(aes2016)$question\_labels <- names(aes2016)

for(i in 1:ncol(aes2016)){

this\_q <- filter(aes2016\_questions, var\_abb == names(aes2016)[i])

# Sometimes a code like 999 'Skipped' is not present in the data but has already

# been replaced with an NA, so we don't want it in our factor level. So we need

# to find out what answers are actually used in the i'th column

used\_a <- unique(pull(aes2016, i))

# the answer labels for this particular question

these\_a <- aes2016\_code %>%

filter(var\_abb == names(aes2016)[i]) %>%

filter(value %in% used\_a) %>%

arrange(value)

attributes(aes2016)$question\_labels[i] <- pull(this\_q, question\_label)

# half a dozen open text questions don't match to the data dictionary so we exclude those:

if(nrow(these\_a) > 0 &

!pull(this\_q, question\_label) %in% c("What kind of work do you do? Full job title",

"What are (or were) the main tasks that you usually perform(ed)?",

"What kind of business or industry is (or was) that in?",

"What was your partner's main activity last week? (other specify)",

"What kind of work does (or did) your partner do? Provide job title",

"Generally speaking, does your partner usually think of himself or herself as Liberal, Labor, National or what? (other specify)")

){

# for the rest, we turn the response into a factor with the correct labels

aes2016[ , i] <- factor(pull(aes2016, i), labels = pull(these\_a, answer\_label))

}

}

aes2016 <- aes2016 %>%

mutate(age\_2016 = 2016 - H2,

agegrp\_2016 = cut(age\_2016,

breaks = c(0, 17.5, 24.5, 34.5, 44.5, 54.5, 64.5, Inf),

labels = c("0-17", "18-24", "25-34", "35-44", "45-54", "55-64", "65+")),

agegrp\_2016 = as.factor(replace\_na(as.character(agegrp\_2016), "Age unknown")),

sex = replace\_na(as.character(H1), "Sex unknown"),

first\_pref\_hr\_grp = case\_when(

B9\_1 %in% c("Liberal Party", "National (Country) Party") ~ "Coalition",

B9\_1 == "Labor Party (ALP)" ~ "Labor",

B9\_1 == "Greens" ~ "Greens",

B9\_1 == "Voted informal" | is.na(B9\_1) ~ "Did not vote/voted informal",

TRUE ~ "Other"

),

first\_pref\_sen\_grp = case\_when(

B9\_2 %in% c("Liberal Party", "National (Country) Party") ~ "Coalition",

B9\_2 == "Labor Party (ALP)" ~ "Labor",

B9\_2 == "Greens" ~ "Greens",

B9\_2 == "Voted informal" | is.na(B9\_2) ~ "Did not vote/voted informal",

TRUE ~ "Other"

),

first\_pref\_sen\_grp2 = case\_when(

B9\_2 == "Liberal Party" ~ "Liberal",

B9\_2 == "National (Country) Party" ~ "National",

B9\_2 == "One Nation" ~ "One Nation",

B9\_2 == "Labor Party (ALP)" ~ "Labor",

B9\_2 == "Greens" ~ "Greens",

TRUE ~ "Other (including did not vote)"

)

) %>%

mutate(first\_pref\_sen\_grp2 = toupper(first\_pref\_sen\_grp2),

first\_pref\_sen\_grp2 = ifelse(grepl("^OTHER", first\_pref\_sen\_grp2),

"OTHER (incl. did not vote)",

first\_pref\_sen\_grp2)) %>%

mutate(first\_pref\_sen\_grp2 = fct\_relevel(first\_pref\_sen\_grp2,

toupper(c("Greens", "Labor", "Liberal",

"National", "One Nation")))) %>%

mutate(tpp\_alp = B11\_1 %in% "Labor Party (ALP)" | B9\_1 %in% "Labor Party (ALP)")

The last few lines of that code also defines some summarised variables that lump various parties or non-votes together and will be useful in later analysis.

**Survey design and weights**

The AES in 2016 used two separate sampling frames: one provided by the [Australian Electoral Commission](https://www.aec.gov.au/), and one from the [Geo-coded National Address File or GNAF](https://data.gov.au/dataset/ds-dga-19432f89-dc3a-4ef3-b943-5326ef1dbecc/details?q=), a large, authoritative open data source of every address in Australia. For inference about Australian adults on the election roll (enrolment is required of all resident adult citizens by legislation), the AES team recommend using a combination of the two samples, and weights are provided in the wt\_enrol column to do so.

The [technical report](https://dataverse.ada.edu.au/file.xhtml?fileId=151&version=2.1) describes data collection as mixed-mode (hard copy and online versions), with a solid system of managing contacts, reminders and non-returns.

The AEC sample is stratified by state, and the GNAF sample by Greater Capital City (GCC) statistical areas (which divide states into two eg “Greater Sydney” and “Rest of New South Wales”). As far as I can see no GCC variable is provided for end-use analysts such as myself. Nor are the various postcode variables populated, and I can’t see a variable for electorate/division (presumably for statistical disclosure control purposes) which is a shame as it would be an important variable to use for a variety of multi-level analyses. So we have to use “state” as our best regional variable.

Post-stratification weighting (raking) was used by the AES to match the population proportions by sex, age group, state, and first preference vote in the House of Representatives. This latter calibration is particularly important to ensure that analysis gives appropriate weight to the supporters of the various parties; it is highly likely that non-response is related to political views (or non-views).

The weights have been scaled to add up to the sample size, rather than to population numbers (as would be the approach of eg a national statistical office). This is presumably out of deference for SPSS’ behaviour of treating survey weights as frequencies, which leads to appalling results if they add up to anything other than the sample size (and still sub-optimal results even then).

Table 16 of the technical report is meant to provide the weighting benchmarks for the post-stratification weighting, but the state, age and sex totals given are only for the GNAF sample; and the benchmarks given for which party voted for appear to be simply incorrect in a way I cannot determine. They add up to much more than either the AEC or GNAF sample, but much less than their combined total; and I can’t match their results with any of the plausible combinations of filters I’ve tried. I’d be happy for anyone to point out if I’m doing something wrong, but my working hypothesis is that Table 16 simply contains some mistakes, whereas the weights in the actual data are correct.

The code below explores the three sets of weights provided, and defines a survey design object (using Thomas Lumley’s survey package) that treats state as the strata for sampling, and sex, state, age and House of Representatives vote as post-stratification variables. Although I won’t make much use of the survey package today, I’ll almost certainly want it for more sophisticated analysis in a later post.

*Post continues after code extract*

#--------------------Weighting-----------------------

# The sample has two parts - addresses supplied by the AEC, and addresses from

# the GNAF (Geocoded National Address File). There are separate weights for each,

# so they can be analysed as two separate surveys or you can use the combined set of weights:

table(aes2016$wt\_aec > 0, useNA = 'always') # for comparison to previous waves of AES

table(aes2016$wt\_gnaf > 0, useNA = 'always') # inference about Australian adults

table(aes2016$wt\_enrol > 0, useNA = 'always') # inference about enrolled Australian adults

# There are 107 cases from the GNAF sample that have zero weight in the final combined sample:

filter(aes2016, wt\_enrol <= 0) %>% select(wt\_aec, wt\_gnaf, wt\_enrol)

# wt\_enrol had been raked for age group, sex, state and first preference vote. See pages 32-33

# of the technical guide.

# These three state variables seem identical, and are all different from those in Table 16

# of the technical report:

cbind(

table(aes2016$finSTATE),

table(aes2016$StateMap),

table(aes2016$SamSTATE)

)

# In fact, Table 16 has the state, age and sex only of the GNAF sample, not the whole sample:

table(filter(aes2016, wt\_gnaf >0)$finSTATE)

table(filter(aes2016, wt\_gnaf >0)$agegrp\_2016)

table(filter(aes2016, wt\_gnaf >0)$H1)

# The first pref party vote in Table 16 looks to be just wrong. The results below match those

# in the code book, but not in Table 16

table(aes2016$first\_pref\_hr\_grp, useNA = 'always')

table(aes2016$first\_pref\_sen\_grp, useNA = 'always')

# Well, we'll assume that wt\_enrol has actually been weighted as described, and ignore

# the problems in Table 16.

# Set up survey design:

aes2016\_svy <- svydesign(~1, strata = ~SamSTATE, weights = ~wt\_enrol, data = aes2016)

# Deduce population totals from the sums of weights (if weights were raked to match pop totals,

# then the sum of the weights should reveal what they were)

age\_pop <- aes2016 %>%

group\_by(agegrp\_2016) %>%

summarise(Freq = sum(wt\_enrol))

sex\_pop <- aes2016 %>%

group\_by(sex) %>%

summarise(Freq = sum(wt\_enrol))

state\_pop <- aes2016 %>%

group\_by(SamSTATE) %>%

summarise(Freq = sum(wt\_enrol))

age2016\_svy <- rake(aes2016\_svy, sample.margins = list(~sex, ~agegrp\_2016, ~SamSTATE),

population.margins = list(sex\_pop, age\_pop, state\_pop))

# Compare the weighted and unweighted response to a survey question

filter(aes2016\_questions, var\_abb == "H14\_2")

svytable(~H14\_2, aes2016\_svy)

table(aes2016$H14\_2)

# Many more "yes" posted answers to the Internet when weighted. So people active on the internet

# needed extra weight (probably younger)

**Analysis – vote and one question**

Now we’re ready for some exploratory analysis. In Australia, the Senate is elected by proportional representation, a wider range of small parties stand candidates, and the first preference vote for that house is less subject to gaming by voters than is the single-transferrable-vote system in the House of Representatives. So I prefer to use Senate vote when examining the range of political opinion, particularly of supporters of the smaller parties. Here is one of the graphics that I love during exploration but which I despair of using for communication purposes – a mosaic plot, which beautifully visualises a contingency table of counts (or, in this case, weighted counts). This particular graphic shows the views of people who voted for different parties in the Senate on the trade off between reducing taxes and spening more on social services:

As any casual observer of Australian politics would have predicted, disproportionately large numbers (coloured blue to show the positive discrepency from the no-interaction null hypothesis) of Greens voters (and to a less extent Labor) are mildly or strongly in favour of spending more on social services. Correspondingly these parties’ supporters have relatively few people counted in the top left corner (coloured red to show a negative discrepancy from the null hypothesis) supporting reducing tax as the priority.

In contrast, there are relatively few Liberal voters in the social services camp, and disproportionately high numbers who favour reducing taxes.

The “other” voters – which includes those who did not vote, could not remember, or did not vote for one of the five most prominent parties – are disproportionately likely to sit on the fence or not have a view on the trade-off between social services and tax cuts, which again is consistent with expectations.

What about a different angle – what influences people when they decide their vote?

We see that Greens voters decided their vote disproportionately on the basis of “the policy issues”. Liberal voters have less people in the “policy” box than would be expected under the no-interaction null hypothesis and were more likely to decide based on either “the party leaders” or “the parties taken as a whole”. National voters stand out as the only party with a noticeable disproportionate count (high, in their case) in the “candidates in your electorate” box, emphasising the known spatially-specific nature of National’s support.

Here’s the code for drawing those mosaic plots, using the vcd R package that implements some useful ways of visualising categorical data.

*Post continues after code extract*

#======================selected bivariate============

#------------------tax/social services-------------

x <- svytable(~E1 + first\_pref\_sen\_grp2, aes2016\_svy)

x <- round(x, 2)

colnames(x)[grepl("OTHER", colnames(x))] <- "OTHER"

rownames(x) <- paste0(str\_wrap(rownames(x), 20))

mosaic(x, shade = TRUE, border = "grey70", direction = "v",

legend = legend\_resbased(10, fontfamily = "Roboto", pvalue = FALSE),

xlab = "x", ylab = "y",

labeling = labeling\_values(

suppress = c(0, 1000),

gp\_labels = gpar(fontfamily = "Roboto", cex = 0.8),

gp\_varnames = gpar(col = "transparent"),

just\_labels = c("center", "right"),

alternate\_labels = c(TRUE, FALSE),

rot\_labels = c(0,90,0,0),

offset\_labels = c(1,0,1.5,0),

digits = 0

))

#--------------how decide vote------------

x <- svytable(~B5 + first\_pref\_sen\_grp2, aes2016\_svy)

x <- round(x, 2)

colnames(x)[grepl("OTHER", colnames(x))] <- "OTHER"

rownames(x) <- paste0(str\_wrap(rownames(x), 20))

mosaic(x, shade = TRUE, border = "grey70", direction = "v",

legend = legend\_resbased(10, fontfamily = "Roboto", pvalue = FALSE),

xlab = "x", ylab = "y",

labeling = labeling\_values(

suppress = c(0, 1000),

gp\_labels = gpar(fontfamily = "Roboto", cex = 0.8),

gp\_varnames = gpar(col = "transparent"),

just\_labels = c("center", "right"),

alternate\_labels = c(TRUE, FALSE),

rot\_labels = c(0,90,0,0),

offset\_labels = c(1,0,1.5,0),

digits = 0

))

**Analysis – vote and a battery of questions**

Like many social sciences surveys, the AES contains a number of batteries of questions with similar response sets. An advantage of this, both for an analyst and the consumer of their results, is that we can set up a common approach to comparing the responses to those questions. Here is a selection of diverging likert plots depicting interesting results from these questions:

Above, we see that Greens voters disproportionately think that Australia hasn’t gone far enough in allowing more migrants into the country, and One Nation voters unsurprisingly tend to think the opposite. Views on change in Australia relating to Aboriginal assistance and land rights follow a similar pattern. Interestingly, there is less partisan difference on some other issues such as “equal opportunities for women”. Overall, the pattern is clear of increasing tendency to think that “things have gone to far” as we move across the spectrum (Greens-Labor-Liberal-National-One Nation).

When it comes to attitudes on the left-right economic issues, we see subtle but expected party differences (chart above). Liberal supporters are particularly inclined to think trade unions have too much power and should be regulated; but to disagree with the statement that income and wealth should be distributed towards ordinary working people.

In the social issues chart above, we see the Greens’ voters strongly disagreeing with turning back boats of asylum seekers, and reintroducing the death penalty – issues where One Nation voters tend to agree with the statement. Again, the clear sequencing of parties’ supporters (Greens-Labor-Liberal-National-One Nation) is most obvious on these social/cultural touchpoint issues, in contrast to the left-right economic debate.

In the above, we see high levels of confidence in the armed forces, police and universities across all parties’ supporters, and distrust in the press, television (what does it even mean to trust the television?) and political parties. Differences by party in trust in institutions varies along predictable party lines eg voters of the Liberal party tend to have less confidence in unions. One Nation voters have particularly low trust in the Australian political system, consistent with the mainstream interpretation in political science of the rise of socially conservative populist parties of this sort.

Attitudes to public spend are fairly consistent across parties, with subtle but unsurprising differences in a few value-based touch points such as defence (Greens favour spending less) and unemployment benefits (Greens favour spending more but the three conservative parties favour spending less; Labor voters are in the middle).

Finally, when it comes to factual knowledge of election law and constitutional issues, there is little to say from a party perspective. Many people don’t know the maximum period between federal elections for the lower house (the correct answer is three); and doubtless niether know nor care about the need to pay a deposit to stand for election. But there is no obvious difference by party voted for.

In summary, on many of the key social variables of the day, we see the expected greatest contrast between supporters of the Greens on one hand and One Nation on the other, with the ALP, Liberal and National voters strung out between. Economic issues indicate another but closely related dimension, with One Nation closer to the Greens than they are to the other socially conservative parties on statements such as “big business in this country has too much power”, “the government should take measures to reduce differences in income levels” and “trade unions in this country have too much power”.

Here’s the code for drawing those diverging-Likert plots. I’m not particularly proud of it, it’s a bit repetitive (I hadn’t realised I’d be doing six of these), but it get’s the job done. There are some specialist R packages for drawing these charts, but I prefer (at this point) build them with the standard grammar of graphics tools in ggplot2 rather learn a new specific API.

* Post continues after code extract\*

#=================Batteries of similar questions========================

#------------------Exploration---------------------

svytable(~ F7 + first\_pref\_sen\_grp, aes2016\_svy, Ntotal = 1000, round = TRUE) # Immigration

svytable(~ F6\_4 + first\_pref\_sen\_grp, aes2016\_svy, Ntotal = 1000, round = TRUE) # war on terror

svytable(~ F2 + first\_pref\_sen\_grp, aes2016\_svy, Ntotal = 1000, round = TRUE) # Republic

svytable(~ E9\_13 + first\_pref\_sen\_grp, aes2016\_svy, Ntotal = 1000, round = TRUE) # trust universities

svytable(~ E9\_14 + first\_pref\_sen\_grp, aes2016\_svy, Ntotal = 1000, round = TRUE) # trust political system

#-----------Confidence in institutions----------------------

d1 <- aes2016 %>%

select(E9\_1:E9\_14, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(institution = gsub("How much .+\\? ", "", question\_label)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, institution) %>%

mutate(negative = value %in% c("Not very much confidence", "None at all"),

prop = freq/sum(freq) \* ifelse(negative, -1, 1)) %>%

mutate(value = factor(value, levels = c("None at all",

"Not very much confidence",

"A great deal of confidence",

"Quite a lot of confidence"))) %>%

ungroup() %>%

mutate(institution = fct\_reorder(institution, prop, .fun = sum))

pal <- brewer.pal(4, "RdYlGn")

names(pal) <- levels(d1$value)[c(1,2,4,3)]

d1 %>%

ggplot(aes(x = institution, fill = value, weight = prop)) +

geom\_bar(data = filter(d1, negative), position = "stack") +

geom\_bar(data = filter(d1, !negative), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = c("None at all",

"Not very much confidence",

"Quite a lot of confidence",

"A great deal of confidence")) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

expand\_limits(y = c(-1, 1)) +

theme(panel.spacing = unit(1.5, "lines"))+

ggtitle("Attitudes to institutions, by first preference Senate vote in 2016",

"How much confidence do you have in the following organisation? ...") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

#-----------more or less expenditure than now----------------------

d2 <- aes2016 %>%

select(D8\_1:D8\_10, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(item = gsub("Should there be .+\\? ", "", question\_label)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, item) %>%

mutate(negative = value %in% c("Much less than now", "Somewhat less than now"),

positive = value %in% c("Much more than now", "Somewhat more than now"),

same = !negative & !positive,

prop = freq/sum(freq) \* case\_when(

negative ~ -1,

positive ~ 1,

TRUE ~ 0.5)) %>%

mutate(value = factor(value, levels = c("Much less than now",

"Somewhat less than now",

"Much more than now",

"Somewhat more than now",

"The same as now"))) %>%

ungroup() %>%

mutate(item = fct\_reorder(item, prop, .fun = sum))

pal <- brewer.pal(5, "RdYlGn")

names(pal) <- levels(d2$value)[c(1,2,5,4,3)]

d2a <- d2 %>%

filter(negative | same) %>%

mutate(prop = -abs(prop))

d2 %>%

ggplot(aes(x = item, fill = value, weight = prop)) +

geom\_bar(data = d2a, position = "stack") +

geom\_bar(data = filter(d2,positive | same), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = levels(d2$value)[c(1,2,5,4,3)]) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

expand\_limits(y = c(-1, 1)) +

theme(panel.spacing = unit(1.5, "lines"))+

ggtitle("Attitudes to public spend, by first preference Senate vote in 2016",

"Should there be more or less public expenditure in the following area? ...") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

#----------change gone far enough----------

d3 <- aes2016 %>%

select(E2\_1:E2\_7, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(item = gsub("Do you think .+\\? ", "", question\_label)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, item) %>%

mutate(negative = value %in% c("Gone much too far", "Gone too far"),

positive = value %in% c("Not gone nearly far enough", "Not gone far enough"),

same = !negative & !positive,

prop = freq/sum(freq) \* case\_when(

negative ~ -1,

positive ~ 1,

TRUE ~ 0.5)) %>%

mutate(value = factor(value, levels = c("Gone much too far",

"Gone too far",

"Not gone nearly far enough",

"Not gone far enough",

"About right"))) %>%

ungroup() %>%

mutate(item = fct\_reorder(item, prop, .fun = sum))

pal <- brewer.pal(5, "RdYlGn")

names(pal) <- levels(d3$value)[c(1,2,5,4,3)]

d3a <- d3 %>%

filter(negative | same) %>%

mutate(prop = -abs(prop))

d3 %>%

ggplot(aes(x = item, fill = value, weight = prop)) +

geom\_bar(data = d3a, position = "stack") +

geom\_bar(data = filter(d3, positive | same), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = levels(d3$value)[c(1,2,5,4,3)]) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

theme(panel.spacing = unit(1.5, "lines")) +

expand\_limits(y = c(-1, 1)) +

ggtitle("Attitudes to change, by first preference Senate vote in 2016",

"Do you think the following change that has been happening in Australia over the years has gone...?") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

#----------agree with various economic statements----------

d4 <- aes2016 %>%

select(D13\_1:D13\_6, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(item = gsub("Do you strongly .+\\? ", "", question\_label)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, item) %>%

mutate(negative = value %in% c("Disagree", "Strongly disagree"),

positive = value %in% c("Agree", "Strongly agree"),

same = !negative & !positive,

prop = freq/sum(freq) \* case\_when(

negative ~ -1,

positive ~ 1,

TRUE ~ 0.5)) %>%

mutate(value = factor(value, levels = c("Strongly disagree",

"Disagree",

"Strongly agree",

"Agree",

"Neither agree nor disagree"))) %>%

ungroup() %>%

mutate(item = fct\_reorder(item, prop, .fun = sum))

pal <- brewer.pal(5, "RdYlGn")

names(pal) <- levels(d4$value)[c(1,2,5,4,3)]

d4a <- d4 %>%

filter(negative | same) %>%

mutate(prop = -abs(prop))

d4 %>%

ggplot(aes(x = item, fill = value, weight = prop)) +

geom\_bar(data = d4a, position = "stack") +

geom\_bar(data = filter(d4, positive | same), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = levels(d4$value)[c(1,2,5,4,3)]) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

theme(panel.spacing = unit(1.5, "lines")) +

expand\_limits(y = c(-1, 1)) +

ggtitle("Attitudes to left-right economic issues, by first preference Senate vote in 2016",

"Do you strongly agree ... or strongly disagree with the following statement?") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

#----------agree with various social statements----------

d5 <- aes2016 %>%

select(E6\_1:E6\_7, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(item = gsub("Do you strongly .+\\? ", "", question\_label)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, item) %>%

mutate(negative = value %in% c("Disagree", "Strongly disagree"),

positive = value %in% c("Agree", "Strongly agree"),

same = !negative & !positive,

prop = freq/sum(freq) \* case\_when(

negative ~ -1,

positive ~ 1,

TRUE ~ 0.5)) %>%

mutate(value = factor(value, levels = c("Strongly disagree",

"Disagree",

"Strongly agree",

"Agree",

"Neither agree nor disagree"))) %>%

ungroup() %>%

mutate(item = fct\_reorder(item, prop, .fun = sum))

pal <- brewer.pal(5, "RdYlGn")

names(pal) <- levels(d5$value)[c(1,2,5,4,3)]

d5a <- d5 %>%

filter(negative | same) %>%

mutate(prop = -abs(prop))

d5 %>%

ggplot(aes(x = item, fill = value, weight = prop)) +

geom\_bar(data = d5a, position = "stack") +

geom\_bar(data = filter(d5, positive | same), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = levels(d5$value)[c(1,2,5,4,3)]) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

theme(panel.spacing = unit(1.5, "lines")) +

expand\_limits(y = c(-1, 1)) +

ggtitle("Attitudes to liberal-conservative social issues, by first preference Senate vote in 2016",

"Do you strongly agree ... or strongly disagree with the following statement?") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

#----------constitutional knowledge----------

d6 <- aes2016 %>%

select(F10\_1:F10\_6, wt\_enrol, first\_pref\_sen\_grp2) %>%

gather(variable, value, -wt\_enrol, -first\_pref\_sen\_grp2) %>%

group\_by(variable, value, first\_pref\_sen\_grp2) %>%

summarise(freq = sum(wt\_enrol)) %>%

ungroup() %>%

left\_join(aes2016\_questions, by = c("variable" = "var\_abb")) %>%

mutate(item = gsub("Do you think .+\\? ", "", question\_label),

item = str\_wrap(item, 50)) %>%

filter(!is.na(value)) %>%

group\_by(variable, first\_pref\_sen\_grp2, item) %>%

mutate(negative = value %in% c("False"),

positive = value %in% c("True"),

same = !negative & !positive,

prop = freq/sum(freq) \* case\_when(

negative ~ -1,

positive ~ 1,

TRUE ~ 0.5)) %>%

mutate(value = factor(value, levels = c("False",

"True",

"Don't know"))) %>%

ungroup() %>%

mutate(item = fct\_reorder(item, prop, .fun = sum))

pal <- brewer.pal(3, "RdYlGn")

names(pal) <- levels(d6$value)[c(1,3,2)]

d6a <- d6 %>%

filter(negative | same) %>%

mutate(prop = -abs(prop))

d6 %>%

ggplot(aes(x = item, fill = value, weight = prop)) +

geom\_bar(data = d6a, position = "stack") +

geom\_bar(data = filter(d6, positive | same), position = "stack") +

facet\_wrap(~first\_pref\_sen\_grp2, ncol = 3) +

coord\_flip() +

scale\_fill\_manual("",

values = pal,

guide = guide\_legend(reverse = FALSE),

breaks = names(pal)) +

scale\_y\_continuous(breaks = c(-1, -0.5, 0, 0.5, 1), labels = c("100%", "50%", "0", "50%", "100%")) +

theme(panel.spacing = unit(1.5, "lines")) +

expand\_limits(y = c(-1, 1)) +

ggtitle("Knowledge of constitutional issues",

"Do you think the following statement is true or false?

(correct answers in order from 'federation in 1901' to '75 members' are True, True, False, False, True, False)") +

labs(x = "", y = "", caption = "Source: Australian Election Study 2016; analysis by freerangestats.info.")

**Final word**

OK, watch this space for more analysis using the AES (2016 and other years), if and when I get time.

Note – I have no affiliation or contact with the Australian Election Study, and the AES researchers bear no responsibility for my analysis or interpretation of their data. Use of the AES data by me is solely at my risk and I indemnify the Australian Data Archive and its host institution, The Australian National University. The Australian Data Archive and its host institution, The Australian National University, are not responsible for the accuracy and completeness of the material supplied. Similarly, if you use my analysis, it is at your risk. Don’t blame me for anything that goes wrong, even if I made a mistake. But it would be nice if you let me know.

thankr::shoulders() %>%

mutate(maintainer = str\_squish(gsub("<.+>", "", maintainer)),

maintainer = ifelse(maintainer == "R-core", "R Core Team", maintainer)) %>%

group\_by(maintainer) %>%

summarise(`Number packages` = sum(no\_packages),

packages = paste(packages, collapse = ", ")) %>%

arrange(desc(`Number packages`)) %>%

knitr::kable() %>%

clipr::write\_clip()