Assignment 2A

Jose Vilardy, Caro Paipa, Jacky Wong, Justin Mah, Seeneyen Youshinee

2 September 2018

# Introduction

There are two types of tables to consolidate for this research:

Table 1: Contains information as Average income, number of individuals, total income by the tax payers, number of individuals with health insurance, number of welfare recipients, total of welfare payment and year.

Table 2: Contains all the type of welfare payments made by the Government each year (Department of Social Services Payment Demographic Data)

The information for each table was structured by year as per the codes below:

# Table 1

### Table 1 Year 2013: Gathering Information for the 2013 Financial Year

library(readxl)  
library(httr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
  
url2 <- "https://data.gov.au/dataset/e29ef9ca-0d1a-47ec-9e9b-14a79a941511/resource/fa6fda82-a70d-4f4c-bb00-203ffc4d20af/download/taxstats2013individual06selected-items-by-taxable-status-postcode.xlsx"  
p2f <- tempfile()  
  
GET(url2, write\_disk(tf2 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/e29ef9ca-0d1a-47ec-9e9b-14a79a941511/resource/fa6fda82-a70d-4f4c-bb00-203ffc4d20af/download/taxstats2013individual06selected-items-by-taxable-status-postcode.xlsx]  
## Date: 2018-09-02 08:02  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 7.46 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file1458fc16433.xlsx

T2013 <- as.data.frame(read\_excel(tf2, sheet = 3, col\_names = TRUE, skip = 2))  
  
  
T2013 <- T2013%>%select("State/Territory1", "Postcode", "Number of individuals\r\nno.", "Number of individuals with private health insurance no.","Taxable income or loss\r\nno.", "Taxable income or loss\r\n$", "Gross interest\r\n$",   
"Australian Government allowances and payments like newstart, youth allowance and austudy payment\r\n$", "Australian Government allowances and payments like newstart, youth allowance and austudy payment\r\nno.")  
  
colnames(T2013) <- c("State", "Postcode", "Population", "NoPeopleHI","NoPeople", "Income", "Interest", "Allowances", "NoAllowances")  
  
rep2013 <- rep("2013", length(T2013$State))  
  
T2013 <- T2013%>%mutate(Year = rep2013)

### Table 1 Year 2014: Gathering Information for the 2014 Financial Year

url3 <- "https://data.gov.au/dataset/25e81c18-2083-4abe-81b6-0f530053c63f/resource/942c9a30-baf2-4432-9513-1395fbef7edf/download/taxstats2014individual06taxablestatuspostcode.xlsx"  
p3f <- tempfile()  
  
GET(url3, write\_disk(tf3 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/25e81c18-2083-4abe-81b6-0f530053c63f/resource/942c9a30-baf2-4432-9513-1395fbef7edf/download/taxstats2014individual06taxablestatuspostcode.xlsx]  
## Date: 2018-09-02 08:02  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 5.02 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file14581f5e77b3.xlsx

T2014 <- as.data.frame(read\_excel(tf3, sheet = 2, col\_names = TRUE, skip = 2))  
  
  
  
  
T2014 <- T2014%>%select("State/Territory1", "Postcode", "Number of individuals", "Number of individuals with private health insurance","Taxable income or loss3\r\nno.", "Taxable income or loss3\r\n$", "Gross interest\r\n$",   
 "Australian government allowances and payments\r\n$", "Australian government allowances and payments\r\nno.")  
  
  
colnames(T2014) <- c("State", "Postcode", "Population", "NoPeopleHI","NoPeople", "Income", "Interest", "Allowances", "NoAllowances")  
  
T2014 <- T2014%>%group\_by(State, Postcode)%>%   
 summarize(Population = sum(Population), NoPeopleHI = sum(NoPeopleHI), NoPeople = sum(NoPeople), Income = sum(Income), Interest = sum(Interest), Allowances = sum(Allowances), NoAllowances = sum(NoAllowances))  
  
  
rep2014 <- rep("2014", length(25))  
  
  
T2014 <- as.data.frame(T2014%>%mutate(Year = rep2014, Postcode = as.character(Postcode)))

### Table 1 Year 2015: Gathering Information for the 2015 Financial Year

url4 <- "https://data.gov.au/dataset/5c99cfed-254d-40a6-af1c-47412b7de6fe/resource/12443334-2273-4e0f-a2e1-c9d0c1d8213b/download/taxstats2015individual06taxablestatusstateterritorypostcode.xlsx"  
p4f <- tempfile()  
  
GET(url4, write\_disk(tf4 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/5c99cfed-254d-40a6-af1c-47412b7de6fe/resource/12443334-2273-4e0f-a2e1-c9d0c1d8213b/download/taxstats2015individual06taxablestatusstateterritorypostcode.xlsx]  
## Date: 2018-09-02 08:03  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 7.72 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file14582f2673dc.xlsx

T2015 <- as.data.frame(read\_excel(tf4, sheet = 3, col\_names = TRUE, skip = 2))  
  
  
  
T2015 <- T2015%>%select("State/ Territory1", "Postcode", "Number of individuals\r\nno.", "People with private health insurance\r\nno.","Taxable income or loss3\r\nno.", "Taxable income or loss3\r\n$", "Gross interest\r\n$",   
 "Australian government allowances and payments\r\n$", "Australian government allowances and payments\r\nno.")  
  
  
colnames(T2015) <- c("State", "Postcode", "Population", "NoPeopleHI","NoPeople", "Income", "Interest", "Allowances", "NoAllowances")  
  
rep2015 <- rep("2015", length(T2015$State))  
  
T2015 <- T2015%>%mutate(Year = rep2015)

### Table 1 Year 2016: Gathering Information for the 2015 Financial Year

url5 <- "https://data.gov.au/dataset/d170213c-4391-4d10-ac24-b0c11768da3f/resource/8c86b6d5-0485-4387-b12b-ad26a9da2033/download/taxstats2016individual06taxablestatusstateterritorypostcodetaxableincome.xlsx"  
p5f <- tempfile()  
  
GET(url5, write\_disk(tf5 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/d170213c-4391-4d10-ac24-b0c11768da3f/resource/8c86b6d5-0485-4387-b12b-ad26a9da2033/download/taxstats2016individual06taxablestatusstateterritorypostcodetaxableincome.xlsx]  
## Date: 2018-09-02 08:03  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 8.36 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file145813a37e8f.xlsx

T2016 <- as.data.frame(read\_excel(tf5, sheet = 3, col\_names = TRUE, skip = 2))  
  
  
  
T2016 <- T2016%>%select("State/ Territory1", "Postcode", "Number of individuals\r\nno.", "People with private health insurance \r\nno.","Taxable income or loss3 \r\nno.", "Taxable income or loss3 \r\n$", "Gross interest \r\n$",   
 "Australian government allowances and payments \r\n$", "Australian government allowances and payments \r\nno.")  
  
  
colnames(T2016) <- c("State", "Postcode", "Population", "NoPeopleHI","NoPeople", "Income", "Interest", "Allowances", "NoAllowances")  
  
rep2016 <- rep("2016", length(T2016$State))  
  
T2016 <- T2016%>%mutate(Year = rep2016)

### Consolidating years 2013, 2014, 2015 and 2016 in Table 1

table1 <- rbind(T2013, T2014, T2015, T2016)  
  
table1 <- table1%>%mutate(AverageIncome = Income/NoPeople)  
  
scores <- quantile(table1$AverageIncome)  
q1 <- quantile(table1$AverageIncome, 0.25)   
q2 <- quantile(table1$AverageIncome, 0.50)  
q3 <- quantile(table1$AverageIncome, 0.75)  
q4 <- quantile(table1$AverageIncome, 1)  
  
table1 <- table1%>% mutate(Q25 = ifelse(AverageIncome <= q1,1,0), Q50 = ifelse(AverageIncome > q1 & AverageIncome <= q2,1,0),   
 Q75 = ifelse(AverageIncome > q2 & AverageIncome <= q3,1,0), Q1 = ifelse(AverageIncome > q3 & AverageIncome <= q4,1,0))

# Table 2

## Table 2 Year 2013: Gathering Information for the 2013 Financial Year

urldss0 <- "https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/6b2343f4-09d3-4b03-a1fe-41897f710c6c/download/dssdemographicsdecember2013.xlsx"  
p7 <- tempfile()  
  
GET(urldss0, write\_disk(tf7 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/6b2343f4-09d3-4b03-a1fe-41897f710c6c/download/dssdemographicsdecember2013.xlsx]  
## Date: 2018-09-02 08:03  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 668 kB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file145816ff6bd6.xlsx

dss2013 <- as.data.frame(read\_excel(tf7, sheet = 7, col\_names = TRUE, skip = 2))  
  
d2013 <- rep("2013", length(dss2013$Postcode))  
  
dss2013 <- dss2013%>%mutate(Year = 2013)  
  
names2013 <- colnames(dss2013)

## Table 2 Year 2014: Gathering Information for the 2014 Financial Year

urldss1 <- "https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/f0615bb3-463f-4352-902c-0b6bb0e22e7d/download/june-2014-dss-demographics-suppressed-2014-final-ftb.xlsx"  
p8 <- tempfile()  
  
GET(urldss1, write\_disk(tf8 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/f0615bb3-463f-4352-902c-0b6bb0e22e7d/download/june-2014-dss-demographics-suppressed-2014-final-ftb.xlsx]  
## Date: 2018-09-02 08:03  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 1.19 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file14585e345d8f.xlsx

dss2014 <- as.data.frame(read\_excel(tf8, sheet = 7, col\_names = TRUE, skip = 2))  
  
d2014 <- rep("2014", length(dss2014$Postcode))  
  
dss2014 <- dss2014%>%mutate(Year = 2014)  
  
colnames(dss2014) <- names2013

## Table 2 Year 2015: Gathering Information for the 2015 Financial Year

urldss2 <- "https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/e6457899-378e-406f-8027-a6ee8a19eec6/download/dss-demographics-june-2015.xlsx"  
p9 <- tempfile()  
  
GET(urldss2, write\_disk(tf9 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/e6457899-378e-406f-8027-a6ee8a19eec6/download/dss-demographics-june-2015.xlsx]  
## Date: 2018-09-02 08:03  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 1.72 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file1458540727b.xlsx

dss2015 <- as.data.frame(read\_excel(tf9, sheet = 9, col\_names = TRUE, skip = 2))  
  
d2015 <- rep("2015", length(dss2015$Postcode))  
  
dss2015 <- dss2015%>%mutate(Year = 2015)  
  
colnames(dss2015) <- names2013  
  
str(dss2015)

## 'data.frame': 2918 obs. of 29 variables:  
## $ Postcode : num 200 800 801 803 804 810 811 812 813 814 ...  
## $ ABSTUDY (Living allowance) : chr "0" "<20" "0" "0" ...  
## $ ABSTUDY (Non-living allowance) : chr "0" "<20" "0" "0" ...  
## $ Age Pension : chr "0" "170" "<20" "<20" ...  
## $ Austudy : chr "<20" "<20" "0" "0" ...  
## $ Carer Allowance : chr "0" "22" "<20" "0" ...  
## $ Carer Allowance (Child Health Care Card only) : chr "0" "0" "0" "0" ...  
## $ Carer Payment : chr "0" "<20" "0" "0" ...  
## $ Commonwealth Seniors Health Card : chr "0" "29" "<20" "0" ...  
## $ Double Orphan Pension : chr "0" "0" "0" "0" ...  
## $ Disability Support Pension : chr "0" "291" "26" "0" ...  
## $ Family Tax Benefit Part A : chr "0" "131" "<20" "<20" ...  
## $ Family Tax Benefit Part B : chr "0" "119" "<20" "<20" ...  
## $ Health Care Card : chr "<20" "321" "<20" "<20" ...  
## $ Low Income Card : chr "<20" "37" "0" "0" ...  
## $ Newstart Allowance : chr "0" "262" "<20" "<20" ...  
## $ Parenting Payment (Partnered) : chr "0" "<20" "0" "0" ...  
## $ Parenting Payment (Single) : chr "0" "<20" "0" "<20" ...  
## $ Partner Allowance : chr "0" "<20" "0" "0" ...  
## $ Pensioner Concession Card : chr "0" "505" "24" "<20" ...  
## $ Sickness Allowance : chr "0" "<20" "0" "0" ...  
## $ Special Benefit : chr "0" "0" "0" "0" ...  
## $ Wife Pension (Partner on Age Pension) : chr "0" "<20" "0" "0" ...  
## $ Wife Pension (Partner on Disability Support Pension): chr "0" "0" "0" "0" ...  
## $ Widow Allowance : chr "0" "0" "0" "0" ...  
## $ Widow B Pension : chr "0" "0" "0" "0" ...  
## $ Youth Allowance (other) : chr "0" "<20" "<20" "0" ...  
## $ Youth Allowance (student and apprentice) : chr "22" "<20" "0" "0" ...  
## $ Year : num 2015 2015 2015 2015 2015 ...

## Table 2 Year 2016: Gathering Information for the 2016 Financial Year

urldss3 <- "https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/829fc910-e263-4a59-ad33-1cd1159ce0df/download/dss-demographics-june-2016.xlsx"  
p10 <- tempfile()  
  
GET(urldss3, write\_disk(tf10 <- tempfile(fileext = ".xlsx")))

## Response [https://data.gov.au/dataset/cff2ae8a-55e4-47db-a66d-e177fe0ac6a0/resource/829fc910-e263-4a59-ad33-1cd1159ce0df/download/dss-demographics-june-2016.xlsx]  
## Date: 2018-09-02 08:04  
## Status: 200  
## Content-Type: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
## Size: 1.7 MB  
## <ON DISK> C:\Users\josev\AppData\Local\Temp\Rtmp2P5mzT\file1458683b5635.xlsx

dss2016 <- as.data.frame(read\_excel(tf10, sheet = 9, col\_names = TRUE, skip = 2))  
  
d2016 <- rep("2016", length(dss2016$Postcode))  
  
dss2016 <- dss2016%>%mutate(Year = d2016)  
  
colnames(dss2016) <- names2013

## Table 2: Consolidation Payment Demographic Table 2

PaymentDemographic<- rbind(dss2013,dss2014, dss2015, dss2016)  
  
PaymentDemographic[PaymentDemographic=="<20"]<-10  
  
PaymentDemographic[is.na(PaymentDemographic)] <- 0  
  
PaymentDemographic[,2:29] <- sapply(PaymentDemographic[, 2:29], as.numeric)  
  
table2 <- PaymentDemographic%>%mutate(Postcode = as.character(Postcode))

## Merging Table1 and Table 2

mergetables <- merge(x= table1, y = table2, by.x = c('Postcode','Year'), by.y = c('Postcode','Year'), all.x = TRUE)  
  
mergetables[is.na(mergetables)] <- 0  
  
write.csv(mergetables, "data\_consolidation.csv")