GovHack – Project Idea

**Award:** <https://2017.hackerspace.govhack.org/award/bounty-plausible-futures>

The ATO is committed to reducing uncertainty about our understanding of the future environment in which we will be operating. Our challenge is for you **to create plausible futures for Australia in 2030** (using the Futures Australia Exec Summary and the APF document, as a primer and guide). Answer this challenge using the ATO Data Sets (which link data from the ATO and Census data from the ABS) along with any other available data.

Remember to explore as many options as possible and explain your assumptions about the trajectory of events and forces at play. Ask yourselves whether the existing data sets present a realistic view of the present, and the extent to which they can be relied upon as a baseline for the study of the future environment.

*Eligibility Criteria*

The ABS and ATO take privacy and confidentiality seriously, so your concept shouldn't attempt to identify any individuals, households or organisations. If you have any questions or concerns about this, please discuss it further with an ABS or an ATO mentor.

*Award Sponsor*

ATO

**Date source:** <http://data.gov.au/dataset/govhackato>

**Team name:** Atomic

**Team members:** Vyanne, Tristan, Jason and Junran

# MVP

* Core model will use the existing ATO-ABS GovHack 2017 dataset and demographic data from the ABS
* Core model will be created in Excel first
* Tristan: may explore graphics (incl. interactive graphs that can be uploaded onto the web) in R’s shiny package
* Vyanne/Jason/Junran: explore the fall-back option of generating the graphics in Excel, perhaps creating a macro to display ‘dynamic charts’ (note that we only need a Proof of Concept to be eligible for this event)

***Central question:***

Australia has enjoyed the mining boom over the last decade, income and wealth inequality has widened during this period. Now that the boom is behind us, innovation plays the key role in driving future productivity growth in both established and emerging industries.

We use the 2006, 2011 and 2016 Census and the ATO data sets published on [data.gov.au](http://www.data.gov.au) in our project to aim to answer the million-dollar question: What would future income and wealth inequality look like for Australia in 2030?

We first attempt to provide a realistic scenario based on an extrapolation of observed demographic trends by age group into the next decade. We hypothesise that income and wealth inequality will be vastly different in income or wealth per capita indicator by 2030. Secondly, we test our hypothesis with different underlying assumptions and visualise every possible outcome.

“Scenario planning is not intended to predict the future, but rather to communicate a wide range of possible outcomes and the consequences of each.”

* Australia 2030 Executive Summary, CSIRO, May 2016

As an extension to the project, we will obtain more comprehensive data to improve our current model. We will devise a self-service website and mobile app that have an interactive interface, which allows the users to select different scenarios with varying demographic assumptions and see how the future will then unfold. For example, a functionality on the user interface will allow the user to select the immigration policy that he/she believes to be most likely and then see its impact on, say, population growth with resultant impacts on income and wealth inequality experienced by remote and rural communities of Australia.

The Government can use our products to assess the impact of their public policies on income and wealth inequality experienced by various groups in the Australian demographics, whether it is rural communities, different age groups, socio-economic status or education levels. General users can reply on our product to assess how their income and wealth track against others from the same city, region or state.

**Further details:**

* Summary of the main features of the ATO dataset
* Unique at the level of ‘Year-Postcode’
* No. of rows = 7350
* No. of years = 3 (2006, 2011, 2015)
* No. of unique postcodes = 2524
* ATO information on: number of taxpayers in a given year-postcode, their total taxable income/loss and associated financial data
* ABS information on: number of people in a given year-postcode classified by (i) age group (ii) ethnicity (iii) gender (iv) self-reported fortnightly income (v) marriage status
* At first glance, we can gain an approximate idea of the following statistics (confining all discussions at the Year-Postcode level for now):
* Growth rate of taxpayers and their total/average taxable income
* Changing demographics according to (i) age group (ii) ethnicity and/or (iii) gender
  + This can shed some light/form the basis of assumptions on mortality, fertility and immigration rates (e.g. according to the ABS data, different ethnic groups seem to have quite contrasting growth rates), which implications on the composition (e.g. age-composition) for the workforce in a decade’s time.
    - Since the composition of the population by age group in each postcode will change at different rates overtime, the proportion of productive workforce will change accordingly.
    - Likewise when we multiply the average income by the number of productive workers over time, the income of the postcode in aggregate will also change.
    - Average income per capita (regardless of age) can be derived by dividing the aggregate income by the total number of residents in the postcode. This can serve as an indicator of financial well-being of all residents.
    - We can then find out if the difference in average income between top performing postcodes and the worst performing to see if that difference widens overtime
  + Note that average income is only applicable to those who can earn an income – i.e. from 15 to 64 year-olds
* As a measure of inequality, we can calculate the Gini Index[[1]](#footnote-2) based on the financial information available (and plot the Lorenz curve! An interactive/dynamic Lorenz curve may be our biggest contribution).

As a ‘first-cut’ of how we can calculate income inequality in a given Year-Postcode :

1. (Divide column E by column D) Calculate
2. Rank in ascending order with
3. Calculate Gini coefficient

* In addition to forecasts at the Year-Postcode level over time, we can also construct graphs based on things such as comparison of a local trend (e.g. Gini index over time) vs trends at the State or National level.
* Some thoughts on the relevance of this topic:
* Changing patterns of income & wealth inequality is highly topical (see all the publicity surrounding Piketty’s *Capital in the Twenty-first century* (2013) as an example)
* Relatively fewer studies have been conducted in Australia[[2]](#footnote-3) in this area to date. However, with the recent debates about housing affordability in metropolitan regions, shifting nature of the workplace from agriculture/manufacturing to innovation (see WA Government’s initiatives for example), an examination of income and wealth inequality over time is quite relevant.
* Feasibility of project complexity and scope:
* At a minimum, the skeleton of the core model (in Excel) should be within our capability to complete by the end of today.
* If we proceed ahead of schedule (TBC at 4:30pm), we can consider: (a) merging with other dataset/s, (b) create an interactive website (a prototype) and so on.

1. We can use competing indices if they are more reflective of the Australian context and/or easier to compute. In the meantime, I will use the Gini index which is the most commonly used indicator and skip over concerns of its limitations. [↑](#footnote-ref-2)
2. In so far as I am aware and certainly not extensive as compared to the voluminous headline grabbing studies of this kind in the United States led by the likes of E. Saez, T. Atkinson as well as columns in the NYT by P. Krugman in the past five years or so. [↑](#footnote-ref-3)