Test

AlistairGJ

06/11/2019

SPELL CHECK

1 Table of Contents

1. Background
2. Data Preprocessing & Visualisation 2.1 Importing & Preprocessing the Activities Meta Data
3. Third Example
4. Fourth Example

CB!!! citation_package: biblatex TABLES + FIGURES CB http://www.ctex.org/documents/packages/ float/caption.pdf https://stackoverflow.com/questions/32634274/knit-hooksset-and-opts-chunkset https: //github.com/yihui/knitr/issues/1102 https://alanarnholt.github.io/GeneralStatistics/rmarkdown/ FormattingTables.html knit_hooks$set(crop = hook_pdfcrop)

```
# eval=FALSE,
import numpy as np
import pandas as pd
import json
PATH = '/Users/alistairgj/Documents/GitHub/IoT_ResearchProject/IoT_November'
```

2 Background

2.1 Introduction

2.1.1 Mankind, Technology & Development

Since the inception of the first home computers in the late 1970's (Press, 1993), modern society has become utterly dependent on and indeed, inexorably bound to digital technology. The rapid and widespread adoption of computational technology has led to the fastest rate of societal and economic development our species has ever experienced. One of the most salient manifestations of technical progress has been the widespread availability and adoption of Information and Communications Technology (ICT), including the rise of the global network of networks known as the Internet.
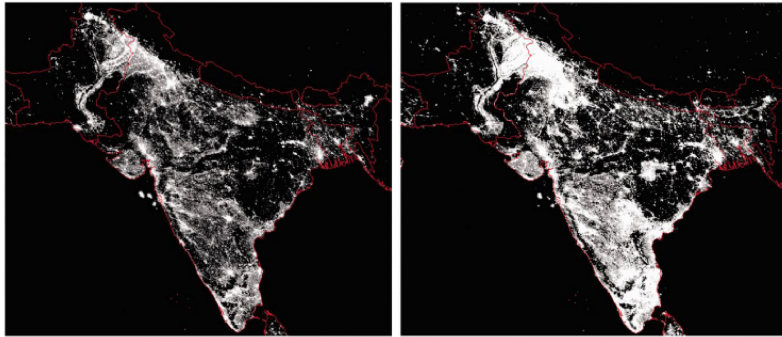

Figure 1: Satellite images of South Asia by night. Left (South Asia in 1994) Right (South Asia in 2010). Images are taken from Maxim Pinkovskiy and Xavier Sala-i-Martin (2016) - Lights, Camera ... Income! Illuminating the National Accounts-Household Surveys Debate. The Quarterly Journal of Economics.
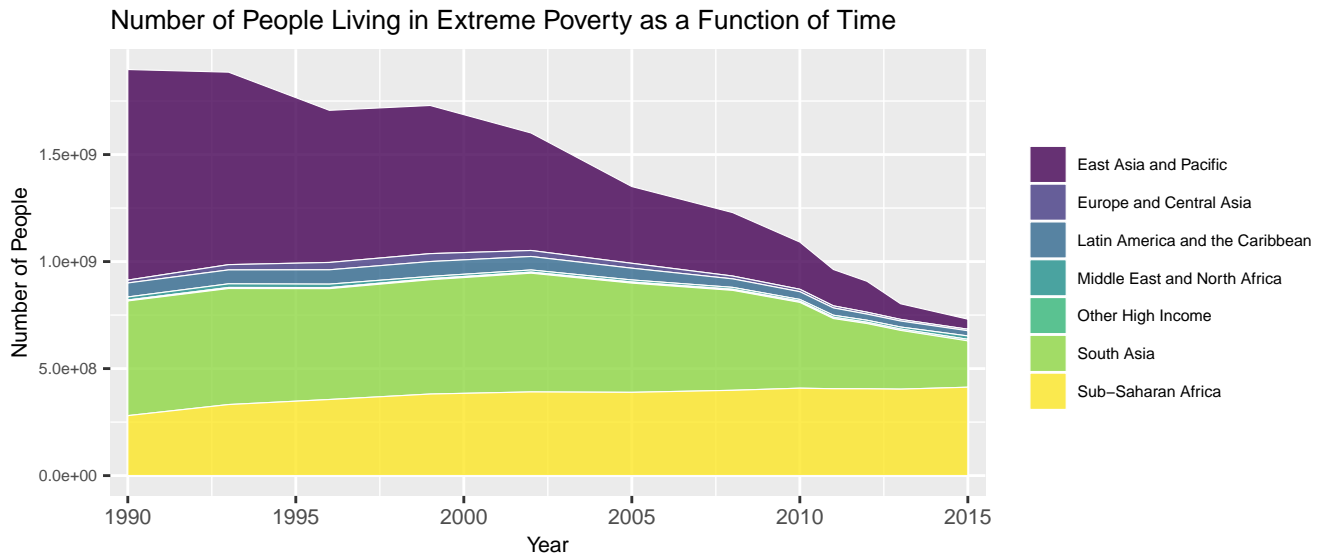

Figure 2: ADD TEXT, Source: Our World in Data

2.1.2 The Rise and Rise of the World Wide Web

According to the International Telecommunication Union (ITU) 2015 ICTs figures, Internet penetration has grown from just over 400 millions users (6 per cent of global population) in 2000 to 3.2 billion users in 2015 (43 per cent of global population), which includes around 2 billion users from developing

countries (Dutta et al., 2015). ICTs bring a broad range of benefits and are recognised as a key to eradicating poverty and unemployment. They enable and facilitate the building a people-centred, inclusive and development-oriented Information Society, where everyone can create, access, utilize and share information and knowledge. This enables individuals, communities and peoples to achieve their full potential in promoting their sustainable development and improving their quality of life ("Information and communication technologies (ICTs) | Poverty Eradication," n.d.).

In addition to a rapidly growing internet user-base both in the developing and developed world, the nature of internet usage has fundamentally changed. Once the purview of academics, engineers and computer scientists sending tiny packets of information back and forth, there are now some 2.5 quintillion bytes of data created each day by all manner of users, industry and sensors to name a few ("Data Never Sleeps 5.0 | Domo," n.d.)
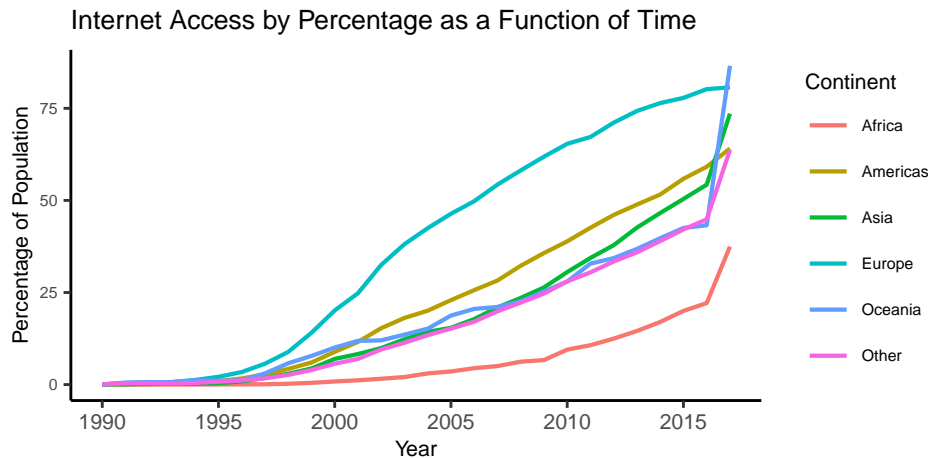

Figure 3: ADD TEXT

2.1.3   Forging a New Technological Paradigm

Such rapid and widespread internet adoption has created a seemingly insatiable demand for exponentially greater computational power and digital storage capacity. This has led to a new and utterly ubiquitous technological paradigm; Cloud Computing. Cloud Computing is succinctly defined as; The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer (https://www.dictionary.com/browse/cloud-computing).

As with the initial rise and widespread implementation of the internet, Cloud Computing itself acts as a facilitator for new technologies. One such example being the Internet of Things (IoT) paradigm. The Internet of Things can be surmised as the extension of the Internet and the Web into the physical realm, by means of the widespread deployment of spatially distributed devices with embedded identification, sensing and/or actuation capabilities (Daniele Miorandi, 2012). The Internet of Things (IoT) paradigm enables physical devices to connect and exchange information, and also allows objects to be sensed or controlled remotely through the internet (Bing Huang, 2018). IoT devices allow objects to be sensed or controlled remotely through the Internet (Luigi Atzori, 2010). IoT thus represents a convergence of real-world objects and digital objects into a unified cyber-physical system.

2.1.4   The Bigger Picture

Considering again the bigger picture of societal benefit, Figure X, shows that the benefits of technological deployment have been widely adopted and exploited. However, the success of modern human society is not

without consequence. All of the benefits our society has enjoyed from the development, production and deployment of technology, has required vast amounts of energy. This energy has, since the industrial revolution, primarily been derived from the burning of fossil fuels (REF). The International Panel on Climate Change (IPCC) notes finds that Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a likely range of 0.8°C to 1.2°C. Global warming is likely to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate (Intergovernmental Panel on Climate Change, 2018).

Climate change poses an existential threat to modern human civilisation, with warming of between 1.5°C and 2°C predicted to cause increases in mean temperature in most land and ocean regions, hot extremes in most inhabited regions, heavy precipitation changes including drought and precipitation deficits in some regions. Additionally, increases in ocean temperature as well as associated increases in ocean acidity and decreases in ocean oxygen levels are projected to reduce risks to marine biodiversity, fisheries, and ecosystems, and their functions and services to humans. Taken together, these effects will lead to risks of the health, livelihoods, food security, water supply, human security, and economic growth of mankind.

Global warming relative to 1850-1900 (°C)

Observed monthly global mean surface temperature

Estimated anthropogenic warming to date and *likely* range

2017

Likely range of modeled responses to stylized pathways

Global CO₂ emissions reach **net zero in 2055** while net non-CO₂ radiative forcing is **reduced after 2030** (grey in **b**, **c** & **d**)

Faster CO₂ reductions (blue in **b** & **c**) result in a **higher probability** of limiting warming to 1.5°C

No reduction of net non-CO₂ radiative forcing (purple in **d**) results in a **lower probability** of limiting warming to 1.5°C
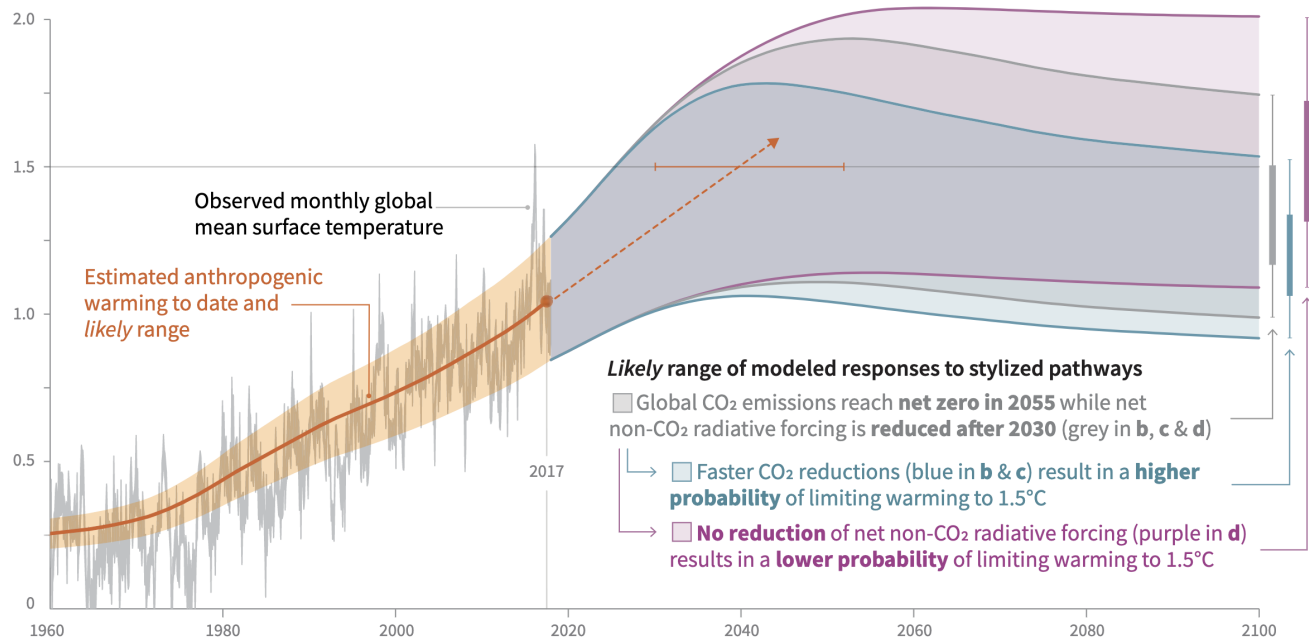
Figure 4: CAPTION

It is therefore imperative moving forward as a species that all steps are be taken to mitigate the emission of greenhouse gases and abate the advance of anthropomorphic climate change. The scale of the challenge is such that technology itself will prove critical in effectively combatting this existential threat to civilisation. When considering energy consumption, the industrial sector (including the non-combusted use of fuels) currently consumes around half of all global energy and feedstock fuels, with residential and commercial buildings (29%) and transport (21%) accounting for the remainder (https://www.bp.com/en/global/corporate/energy-economics/energy-outlook/demand-by-sector.html).

2.1.5 Optimizing Global Energy Usage COMPLETE

With residential and commercial buildings accounting for 29% of energy demand globally (REF). A qualitative assessment of figures X Y and Z show that our increased energy consumption has coincided

with our staggering global achievements, including lifting millions out of poverty blah blah blah. We have more development we use more energy, but we have more development we are better off in all ways. We ALSO have more internet connectivity. It therefore stands that the key to continues human prosperity is to de-couple energy demand growth from economic growth. Clearly economic prosperity is a direct contributing factor to human prosperity in general. One such avenue for reduced energy consumption is the optimization of existing services, utilities, and so on with respect to the timing, manner and BLAH of their energy consumption.

3 Data Preprocessing & Visualisation

3.1 Importing & Preprocessing the Activities Meta Data

The dataset `S1Activities.csv` was imported into the interactive development environment. These data contains a tabulated summary of Heading, Category, Subcategory and a corresponding code. After importation, the dataset has dimensionality of [3, 33], with `Heading`, `Category` & `Subcategory` present as non-null objects, as seen in Table **??** below. The attribute `Code` (which codefies the unique set of Heading, Category) was imported as an index value. At this time, the activities data will not be kept in it's native state and will not be subject to preprocessing.

Table 1: The S1 activities dataset

	Heading	Category	Subcategory
1	Employment related	Employment work at home	Work at home
5	Employment related	Travel employment	Going out to work
10	Personal needs	Eating	Eating
15	Personal needs	Personal hygiene	Toileting
20	Personal needs	Personal hygiene	Bathing

3.2 Importing & Preprocessing the Sensor Meta Data

The dataset `S1sensors.csv` was imported into the interactive development environment. These data contains a tabulated values for Sensor ID, Room and Sensor Activity Type, with no header row present in the original dataset. After importation, the dataset has dimensionality of [3, 76], with header 0, 1 & 2 corresponding to SensorID, Room & Sensor Activity Type, respectively, as seen in Table 2. All attributes are nominal, and were imported as dtype str, accordingly. Attribute attribute 1 & 2 contain degenerate values (e.g., Bathroom & Light Switch). This will be addressed in the subsequent data preprocessing. The preprocessing of the sensor data is a critical step in our analysis. Careful consideration of the data, including the presence of duplicates. This is because if we dont have a sufficient understanding of where and why duplicates exist, we will not be able to satisfactorarily preprocess them. Failure to do so we mean that there is potentail degeneracy in our source dataset, leading to unknown issues with our downstream analysis.

Table 2: The S1 sensor meta data

0	1	2
100	Bathroom	Toilet Flush
101	Bathroom	Light switch
104	Foyer	Light switch
105	Kitchen	Light switch
106	Kitchen	Burner

Column [1] & Column [2] of the sensor data will be concatenated, whitespace will be removed, all text will be cast to lowercase and a final whitespace strip will be performed. The python script `S1sensorsPreprocessing.py` is run perform several preprocessing steps in these data. The script concatenates the attributes `dsS1Sensors[1]` and `dsS1Sensors[2]`, with an underscore. Whitespace is then stripped and all string values are coerced to lowercase. This newly created attribute is then added to the dataframe, as seen below (REF). Additionally, the attributes 0, 1 & 2 are renamed `subActNum`, `room` & `activity`, respectively.

- Data types

- IF a sub-act requires electricity

Table 3: The first iteration of processed S1 sensor meta data

subActNum	room	activity	concat
100	Bathroom	Toilet Flush	bathroom_toiletflush
101	Bathroom	Light switch	bathroom_lightswitch
104	Foyer	Light switch	foyer_lightswitch
105	Kitchen	Light switch	kitchen_lightswitch
106	Kitchen	Burner	kitchen_burner
107	Living room	Light switch	livingroom_lightswitch

The function `getUniqueValues.py` is invoked to provide a means of capturing a list of unique values in a given column of a dataset. The newly created `dsS1Sensors` is then checked for duplicates in two attributes, `subActNum` & `concat`. The number of unique values in `dsS1Sensors.subActNum` is found to be 76, demonstrating that this attribute contains a set of completely unique values (recall n(rows) = 76). The number of unique values in `dsS1Sensors.concat` is found to be 41, demonstrating that despite the concatenation methodology, their are still duplicate values in the dataframe. These duplicate values warrant further investigation. The function `length(unique())` was used to list counts with their corresponding values.

- `length(unique(py$dsS1Sensors$subActNum)) = 76`
- `length(unique(py$dsS1Sensors$concat)) = 41`

COMMENT

Upon compilation of the above summary list, and with reference to the original work (REF) is was determined that these values result from multiple sensors with extremely similar functionality. For example, kitchen_burner has a value of n=4 - this is because on the burner in the apartment under investigation, there were 4 individual burners present. Similarly, kitchen_cabinet has a value of n=15, indicating that for the various cabinets in the apartment, each were given sensors. On the one-hand, this level of granularity may provide fertile grounds for advanced analysis, HOWEVER, for the purposes of this research project, such values will serve to increase the dimensionality of the overall dataset. High dimensionality can lead to difficulties with plotting, ... ML (REF) and thus IN A SUBSEQUENT PREPROCESSING exercise these values will be collapsed down to have n=1.

Creation of JSON Catalogues PRIOR to dup removal - why? Because even if a key-value pair cannot be matched it will simply be ignored **Prior to dupe removal** As this work is largely concerned with energy usage in the home, the sub-activities will be categorized based on their energy requirement. That is, if a sub-activity requires an input of energy beyond what the end user alone can provide, it will be classified as `energyReq` = true. Whereas, if a sub-activity is able to be performed through only interaction with the end user, it will be classified as `energyReq` = false. By way of example, the sub-activity `bathroom_toiletflush` will have an `energyReq` equal to false, while the sub-activity `bathroom_lightswitch` will have an `energyReq` equal to true. Each row (n=76) needs to be inspected manually to determine if the activity requires electricity.

Function `reqEnergy_containSpecialChar.py` is run - After visual inspection, uses `reqEnergy = 'ligh|burn|mach|toas|freez|dvd|lamp|washer|dry|exh|disp|frig|oven|hot|shower|micro'` to find subActivities which require energy input beyond that of the end user - Checked below for SPECIAL CHAR - Special CHARs removed

Table 4: Demo table

	subActNum	room	activity	concat	reqEnergy	specialChar
58	82	Office/study	Drawer	office/study__drawer	FALSE	TRUE
68	92	Office/study	Light switch	office/study__lightswitch	TRUE	TRUE

```
#%run -i reqEnergy_containSpecialCharClean.py
dsS1Sensors['concat'].replace('office/study_drawer','study_drawer',inplace=True)
dsS1Sensors['room'].replace('Office/study','Study',inplace=True)
dsS1Sensors['concat'].replace('office/study_lightswitch','study_lightwitch',inplace=True)
dsS1Sensors['subActNumConcat'] = 'subActNum_' + dsS1Sensors['subActNum'].astype(str)
dsS1Sensors.drop(columns=['specialChar'], inplace=True)
```

Table 5: Demo table

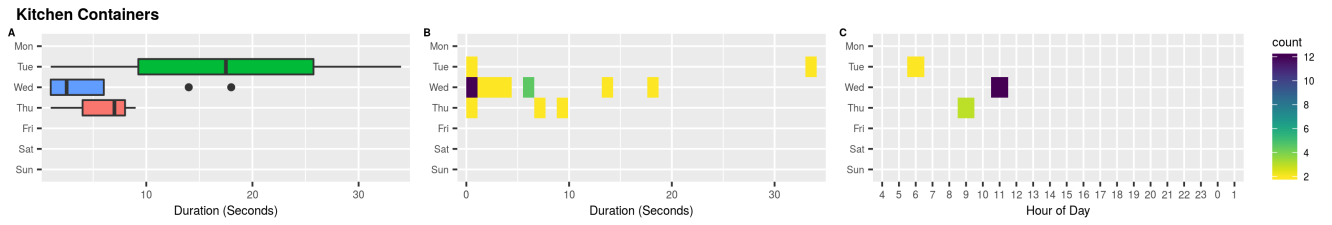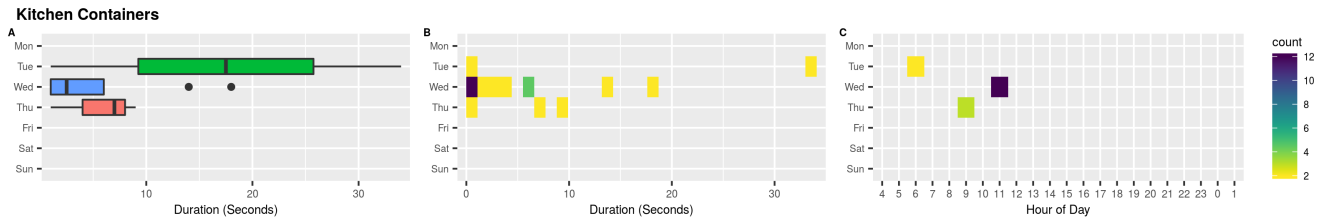|	subActNum	room	activity	concat	reqEnergy	subActNumConcat
59	82	Study	Drawer	study_drawer	FALSE	subActNum__82

Figure 5: A caption

Figure 6: A caption

```
%run -i reqEnergy_containSpecialCharClean.py
```
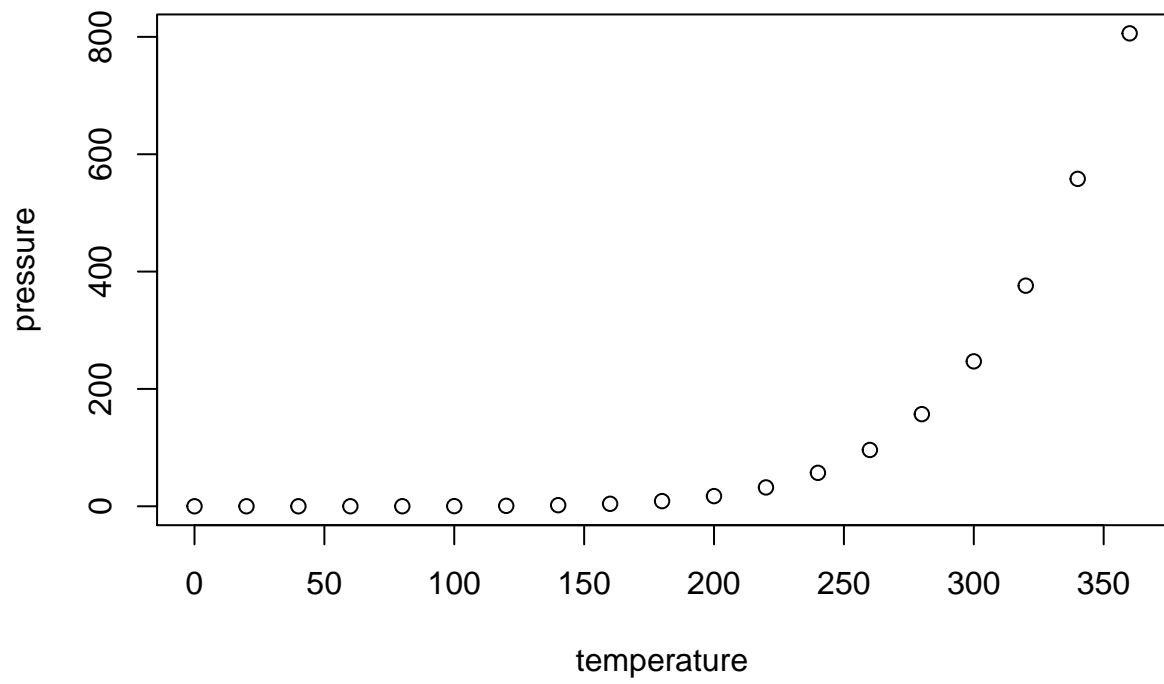
Test for in 7 we see XYZ

Figure 7: This is a caption

Reduce image border (not working???) https://holtzy.github.io/Pimp-my-rmd/