DATA VISUALIZATION REPORT

FIT5147 Final Project

Abstract

A report addressing my findings of Energy Consumption in Australia

Author: Haoheng Zhu (30376467)

Lab 13: Mohammad Haqqani

Table of content

Introduction

My target audience is general audience. My motivation for initiating this project is originally inspired by my own curiosity of finding the energy consumption preference between Renewable and Non-renewable energy change over the past decades in Australia.

After the data exploration, in the final visualization I would like to convey that:

- Throughout the decades, all states follow the same pattern that the greater the GSP, the greater its energy consumption.
- Although in general the energy consumption preference tilts to Renewable energy, in NSW the preference (consumption) of renewable energy increases greatly among all.
- There is, in general, a negative correlation between Renewable and Non-renewable energy consumption. Among all states, "SA, NSW, VIC" have the most obvious negative correlation between Renewable and Non-renewable energy consumption.

Design

In the **brainstorm step**, I thought about using line charts to show correlation for Renewable and Non-renewable energy along with GDP growth by normalizing the datasets, I thought about using grouped bar charts to compare renewable and non-renewable energy for each state, and I really would like to visualize the data on a cartogram map with mini charts.

I thought about showing the [import vs export] and [consumption vs production] of the energy sources and letting audience to have the options to choose specific energy type (oil, coal, solar, wind etc) to visualize. But then I realize this isn't a good idea because too much options would confuse my audience and drive them away from focusing on what message I want to convey. So, I filtered out these ideas.

In **design 1**, I tried to deliver the messages through a line-chart with normalized data for Renewable, Non-renewable energy and GSP. This way, audience can see the negative correlation between Renewable and Non-renewable energy, also a positive correlation between energy consumption and GSP. I would show a default chart by choosing the state that have the most negative correlation between Renewable and Non-renewable energy. Then I let audience to have the options to select from states and discover the same pattern for most (if not all) of the states. Also, after conveying my message, I would like to give audience the options to choose from different energy resources to discover which type of energy has the most negative correlation if they are interested in. But again, I later in the implementation removed most of the options to better let audience focus on the message I want to receive.

The advantages of this design are that it best shows the trend for energy consumption over the years along with GSP. This design is very easy and familiar to read for audience, very clean graph, conveying the very messages with no unnecessary distractions.

The disadvantages of this design are that the audience have to switch between states and can't compare multiple states simultaneously.

In **design 2**, I come up with a continuous cartogram map with mini pie chart and time frame. Size of the state changes according to GSP, pie chart displays the proportion of Renewable and Non-renewable energy consumption, size of pie chart changes according to the total consumption of energy.

By this design, audience are able to compare the changes of energy consumption from different time frame for each state simultaneously, and not only that, but also able to see the total energy consumption (size of the pie) changes according to GSP (size of the state). Audience would see the negative correlation between Renewable and Non-renewable energy by seeing the proportion of Renewable energy gradually enlarges over the years.

The only operation for the audience is to change the time frame they would like to see.

Advantages of this design are that audience are able to see the correlations for all states simultaneously, and they can see them on a map which is very perceptual.

However, the disadvantages are: audience still needs to switch between time frames in order to see the changes over time, also that for some states, the negative correlation between Renewable and Non-renewable energy isn't obvious enough to be identified from a small pie chart.

In **design 3**, I focus on solving the time frame issue by designing a motion chart. By default, on the y-axis it goes with energy consumption, on x-axis it goes with GSP, the circles are the states, and the sizes changes by energy type (Renewable or Non-renewable).

With the motion chart, audience should be able to see that over the years, the circles move from lower left (low GSP low energy consumption) to upper right (high GSP high energy consumption). The size of the circles should change greater when the audience choose Renewable energy rather than Non-renewable energy. This is due to the fact that the consumption of renewable energy grows faster than Non-renewable energy in the extend of growth rate.

I would like to give audience the options to choose their own y-axis and x-axis, as well as size of the bubble. The options of choosing from different source of energy, population, GSP. This way, after the audience went through my default motion chart, they are able to explore other patterns based on their own interest. Basically, I noticed that the energy consumption is also positively correlated to population which makes sense just from intuition.

The advantages from this design are: motion chart enables the observation of multiple variables movement on a time series, it shows energy consumption on the y-axis, GSP growth on the x-axis, type of energy consumption as size, all the states as circles on the same chart simultaneously.

The disadvantages though, are: it could be hard to notice the negative correlation between Renewable energy and Non-renewable energy.

In **final design**, I choose to visualize with cartogram and mini bar chart to show both Renewable energy and Non-renewable energy along with a horizontal line separating import and export. Bars above the horizontal line would be export and bars below the horizontal line would be import. When hover on a specific state, displays a normalized line chart showing the negative correlation between Renewable and Non-renewable energy for that state along with its GSP.

Implementation

When I start to implement my designs, I realized that I could use the motion chart to tell the story of how energy consumption changes along with GSP for all states over the years. Then I use an animated time series map with mini charts showing how the proportion of Renewable and Non-renewable energy change over the years. Finally, with an animated line chart comparing the negative correlation between Renewable and Non-renewable energy in details for each state.

The purpose of motion chart is to give audience a general idea, or an overview of 'energy consumption grows as GSP grows'. Then the map with mini chart serves the purpose of giving the audience a general idea of 'Renewable energy consumption grows faster than Non-renewable energy', Australia has been increasingly consuming more of Renewable energy rather than Non-renewable energy in general. But one problem for this chart is that for some states, it's hard to see this trend on a small size mini chart. The third animated line chart would complement with the second chart in a way that it would emphasis the negative correlation between Renewable and Non-renewable energy by plotting their logarithm values (logarithm returns the change in rate).

To better guide the audience towards coming to my conclusion when exploring the visualization, I put some questions on the top of the pages and hope audience to answer them by playing with the visualization.

Before I can implement the designs, I have to wrangle the datasets so that they are in the appropriate format for the charts.

For **motion chart**, I will need the dataframe to be in long format like this:

State	Year	Population	GSP	Energy.consumption	Energy.consumption.per.capita	Energy.intensity	Energy.productivity
NSW	1989	6116232	290456	1230.400	201.170	4236.098	236.066
NSW	1990	6188051	291736	1228.600	198.544	4211.342	237.454
NSW	1991	6252709	292512	1219.400	195.019	4168.718	239.882
NSW	1992	6294808	300805	1249.100	198.433	4152.524	240.817
NSW	1993	6347013	312254	1276.700	201.150	4088.659	244.579
NSW	1994	6411398	324144	1314.600	205.041	4055.605	246.572
NSW	1995	6486090	337037	1351.200	208.323	4009.055	249.435
NSW	1996	6556800	349531	1380.600	210.560	3949.864	253.173
NSW	1997	6617331	361983	1382.000	208.846	3817.859	261.927

But my datasets are like this:

Table B2
New South Wales - population, GSP and energy consumption

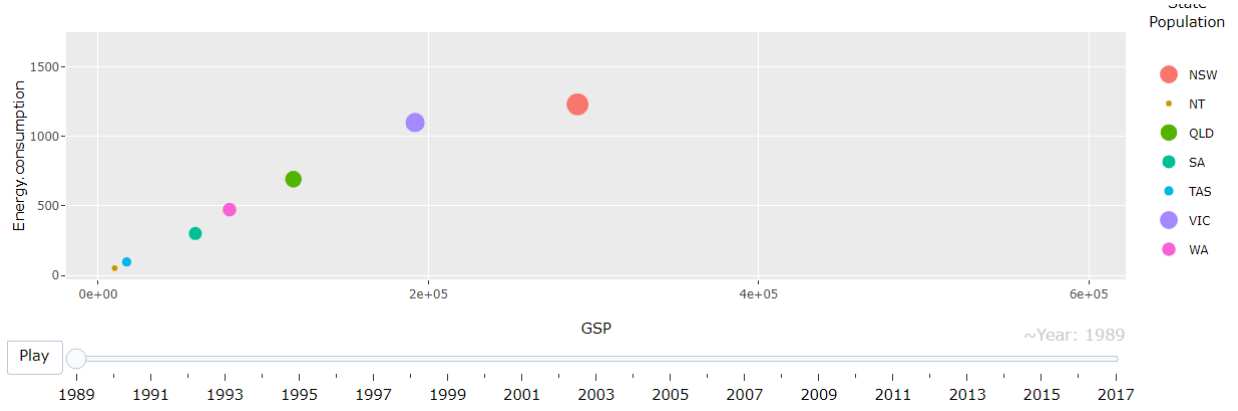
	Population	GSP	Energy consumption	Energy consumption per capita	Energy intensity	Energy productivity
	Number	$ million	PJ	GJ/person	GJ/$ million	$ million/PJ
1960-61	3,977,329		568	143		
1961-62	4,053,121		579	143		
1962-63	4,123,398		595	144		
1963-64	4,188,248		639	152		
1964-65	4,263,903		678	159		
1965-66	4,333,933		702	162		
1966-67	4,398,717		740	168		
1967-68	4,471,420		780	174		
1968-69	4,562,850		810	178		
1969-70	4,653,798		833	179		
1970-71	4,876,672		853	175		
1971-72	4,954,898		884	178		
1972-73	5,015,204		927	185		
1973-74	5,080,294		943	186		
1974-75	5,131,023		965	188		
1975-76	5,167,328		934	181		
1976-77	5,215,576		980	188		
1977-78	5,271,771		1,011	192		
1978-79	5,331,927		1,019	191		
1979-80	5,395,818		1,059	196		

Title page | AUS | NSW | VIC | QLD | WA | SA | TAS | NT

Thus, I have to create a new dataframe and extract the needed data from different excel sheets. To read the excel sheets, since they are in xlsx format, I read them with 'readxl' library and specified which rows, columns, sheets to read from. Then I wrangle the dataset with 'VIM', 'plyr','dplyr', 'forcats', 'tidyverse' libraries. For some columns I have to rename them, for example the Year column, for some columns I have to change the data-type in order to generate time series animation.

After that, I tried to implement motion chart with 'googleVis' and 'ggplot2'. But I found that with 'plotly' library it gives the best and simplest result.

When implementing the motion chart, I realized that I should no longer let audience have the options to choose from different energy sources because now the motion chart should focus on showing how energy consumption changes along with GSP. I still allow audience to choose from Energy consumption, Energy per capita, Energy density, Energy Productivity because they all complement each other. Also, I allow audience to switch between GSP and population for y-axis and size because they are highly positively correlated with energy consumption. All these combinations should come to a very similar conclusion that the energy consumption does grow as GSP/population grows.



From the two screenshots, one would notice that all the circles move towards upper right (high GSP high energy consumption) although 3 states' movements are minimal.

For the **map with mini chart**, I needed the geolocation for each state and total non-renewable, total renewable data. I need the dataframe to be like this:

Year	NonR	Renewable	State	lon	lat
2008-06-19	70456.700	3926.000	NSW	146.9211	-31.25322
2009-06-19	69579.100	5053.600	NSW	146.9211	-31.25322
2010-06-19	65802.400	6936.700	NSW	146.9211	-31.25322
2011-06-19	64707.900	5970.400	NSW	146.9211	-31.25322
2012-06-19	62158.200	8180.600	NSW	146.9211	-31.25322
2013-06-19	61518.495	6970.133	NSW	146.9211	-31.25322
2014-06-19	57374.890	6784.263	NSW	146.9211	-31.25322
2015-06-19	60252.593	9997.627	NSW	146.9211	-31.25322
2016-06-19	59619.497	11388.253	NSW	146.9211	-31.25322
2017-06-19	60403.598	11345.470	NSW	146.9211	-31.25322
2008-06-19	53627.400	1446.900	VIC	144.7852	-37.47131
2009-06-19	53356.600	2619.300	VIC	144.7852	-37.47131

But my datasets are like this:

Table O2
Electricity generation in New South Wales, by

	2008-09	2009-10	2010-11	2011-12
	GWh	GWh	GWh	GWh
Non-renewable fuels				
Black coal	67,650.3	64,398.6	60,732.2	60,453.7
Brown coal				
Natural gas	2,445.1	4,681.3	4,742.7	4,212.1
Oil products	24.7	5.6	61.4	42.1
Other **a**	336.6	493.6	266.1	
Total non-renewable	**70,456.7**	**69,579.1**	**65,802.4**	**64,707.9**
Renewable fuels				
Bagasse, wood **b**	291.3	295.3	253.3	413.6
Biogas **b**	383.4	385.8	360.0	408.7
Wind	41.3	432.5	530.1	697.6
Hydro	3,173.7	3,820.8	5,267.0	3,792.8
Large-scale solar PV				
Small-scale solar PV	36.3	119.2	526.3	657.7
Geothermal				
Total renewable	**3,926.0**	**5,053.6**	**6,936.7**	**5,970.4**
Of which ACT				
Total	**74,382.7**	**74,632.7**	**72,739.1**	**70,678.3**

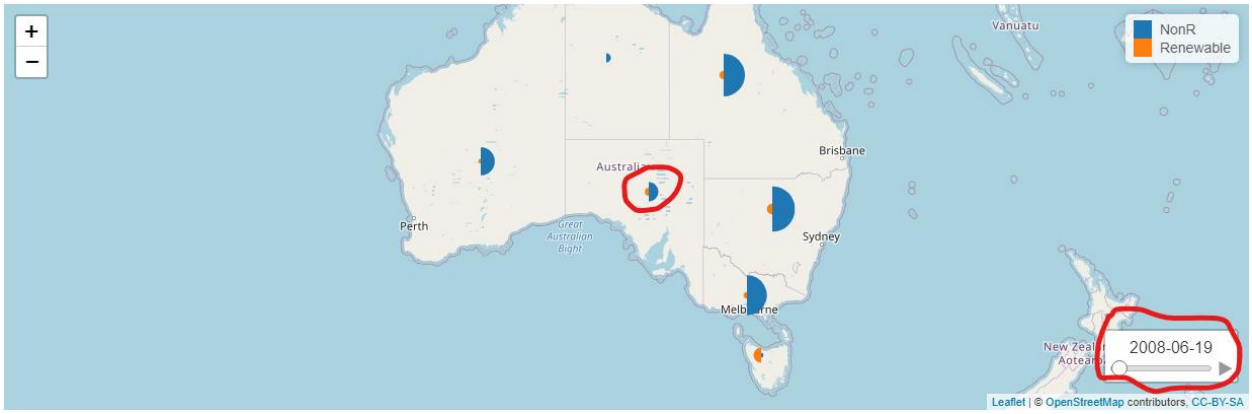
Calendar year

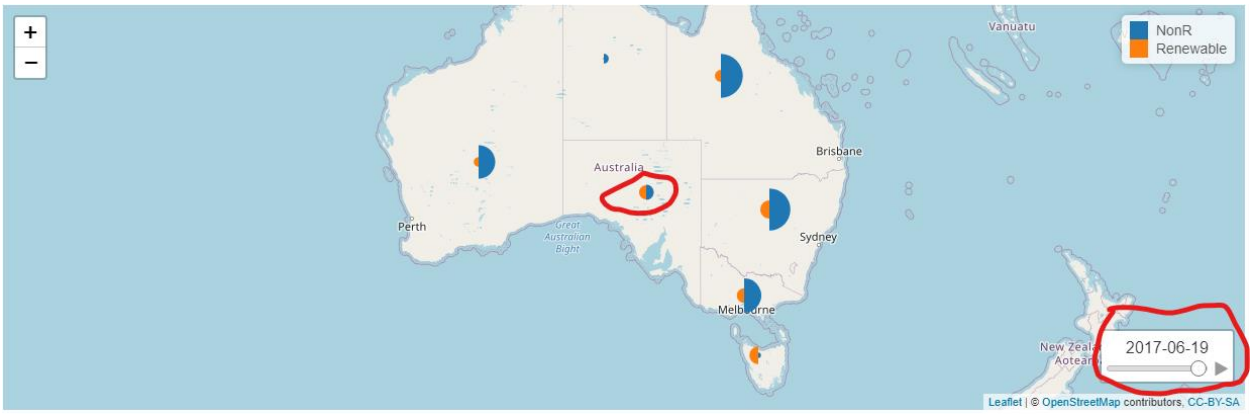
Title page | Index | AUS | NSW | VIC | QLD | WA | SA | TAS | NT

In order to get the geolocation, I used ‘ggmap’ and ‘sp’ to retrieve the longitudes and latitudes for each state. Then I performed similar dataframe transform and reform as I did in the motion chart step for my datasets.

I attempted to create a cartogram with libraries ‘cartogram’, ‘broom’, ‘tweenr’ but the distortion of the states made it very difficult to implement mini charts so I changed to ‘leaflet’.

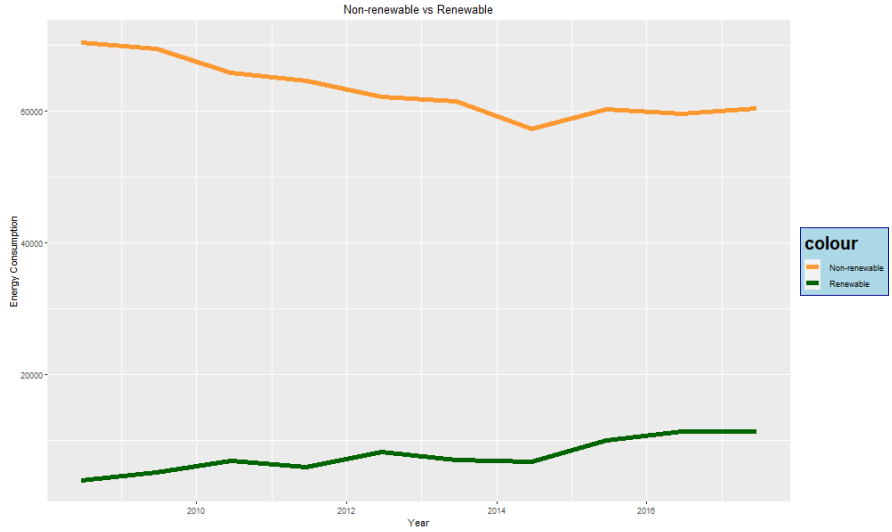
My original approach was to use addMarker function on leaflet map then generate a mini chart on tooltip or popup. But along my way of searching for online examples, I found this powerful library ‘leaflet.minicharts’. It comes with 4 chart types: polar-radius, polar-area, pie, bar which are exactly what I wished to display. By default, I choose the polar-area mini chart because the size of area changes with the total consumption of that type of energy. As shown in the bellowing 2 screenshots, if you look closely on SA, the size of the Renewable (orange) becomes as large as the size of Non-renewable (blue) just in 10 years.



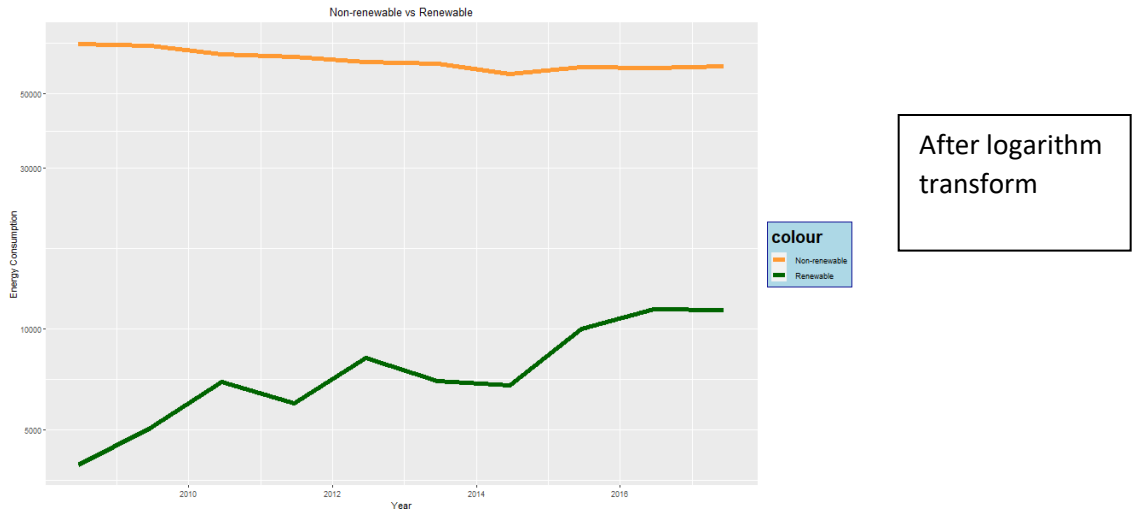


The **polar-area** effect works better than pie chart in this case because although the growth rate for Renewable energy is large, the total consumption of renewable energy is still relatively small comparing with the consumption size of Non-renewable energy. If a mini pie chart is applied here, one would be difficult to spot the proportion change for many states.

For the **animated line chart**, I applied **logarithm transform** on the y-axis so that the change rate is much more obvious than without the transform. This is due to the total size of Renewable and Non-renewable are very different, it would be hard to compare their change on the same small graph. The libraries needed to generate the animation are ‘gganimate’, ‘gifski’, ‘av’ and the plot tool is ‘ggplot2’.



Without logarithm transform



These two screenshots are from NSW but the effect of logarithm plays big. With the log transform, one can way better notice the change rate when comparing in the same graph.

User guide

My visualizations are developed on **Shiny** with R codes.

My submit files contain two r-code files, 2 wrangled csv files, 2 raw data files. In the two r-code files, 'Final project r coding' contains the coding for data wrangling, charts testing; the 'final project shiny' contains the shiny app. The 2 wrangled csv files are 'data_All_state.csv' and 'full_df.csv'. These two csv files are the result dataframe after wrangling the raw datasets. The 2 raw data files are 'australian_energy_statistics_2019_table_b.xlsx' and 'australian_energy_statistics_2019_table_o.xlsx'. These two excel files contain the raw data needed for my final visualizations.

Running the 'final project shiny' code file should be sufficient to run the shiny app.

The basic structure of the app is this:

- Navigation Bar: 4 tab panels (Introduction, Motion Chart, Map & mini charts, Animated line chart)
- Title Panel: Description of the page.
- Side bar: SelectInputs

The **first page** of the shiny app are some texts with Introduction of my motivation for this project and what questions I intend to address.

The **second page** is motion chart with side bar choices of:

- Y-axis: Energy consumption, Energy consumption per capita, Energy intensity, Energy productivity (default setting is Energy consumption)

- X-axis: GSP, Population (default settings is GSP)
- Size: GSP, Population (default settings is Population)

Operation on this page is to choose inputs for these variables then click on play button on the motion chart. After the animation, if click on the graph, a tooltip would pop up with conclusion of my finding for that chart.

The **third page** is Map & mini charts with side bar choices of:

- Chart type: Polar-area, Bar, Pie, Polar-radius (default: Polar-area)

Operation for this page is to choose a mini chart type for the map, then click on play button on the map. After the animation, if click on the graph, a tooltip would pop up with conclusion of my finding for that chart.

The **fourth page** is Animated line chart with side bar choices of:

- State: NSW, TAS, VIC, QLD, WA, SA, NT (default: SA)

Operation for this page is to choose a state, then an animated line chart would be generated in about 15 sec. After the animation, if click on the graph, a tooltip would pop up with conclusion of my finding for that chart.

Conclusion

From this project, I learned to use many new libraries that I can use to generate interactive map as well as animated map and line chart. I found 'plotly' to be extremely powerful when generating interactive charts. It returns a tooltip with detailed data when you hover on the chart. I found that 'ggmap' is very powerful when generating a map with appropriate geospatial data. I found that 'gganimate' is very handy in terms of generating animated charts. The idea behind 'gganimate' is simply generating many frames of png and play a gif animation. It's simple but useful.

There were some difficult times:

- When I couldn't generate the same chart with shiny app using the same codes from my test codes. It took me a long while to figure out that the reason why shiny couldn't generate the charts was because it couldn't read the input data from aes() function. In shiny app, one has to use aes_string() if there is input data from user.
- When I couldn't generate animated line chart with shiny app using the same codes from test codes. It also took me a long while to figure out that the reason was

because shiny couldn't display the animation through 'gifski' and 'av' library. One has to save the gif temporarily and later delete the file.

- When I tried to filter dataframe with user input on shiny app. It took me a lot of searching to find some hints on how to filter dataframe with shiny app.
- When I read the wrangled csv file, the Year column wasn't read as date data type but factor type. It took me a long while to realize this issue and I had to change the data type again after reading the csv file.
- When I wish to find out what data format is needed to generate interactive map, motion chart. It took me a lot of searching and time to find the sample data for the charts.

I'm very happy with all the charts that I come up with. They very well conveyed my messages to the audience and created an easy-to-interact, easy-to-understand visualizations for general audience.

If I were to do all these again, I would try to implement the visualizations with D3.

Bibliography

A Grammar of Animated Graphics (https://github.com/thomasp85/gganimate).

Package 'gganimate' (https://cran.r-project.org/web/packages/gganimate/gganimate.pdf).

Interactive web-based data visualization with R, plotly, and shiny (https://plotly-r.com/).

plotly v4.9.2.1 (https://www.rdocumentation.org/packages/plotly/versions/4.9.2.1).

Package 'ggmap' (https://cran.r-project.org/web/packages/ggmap/ggmap.pdf).

Introduction to leaflet.minicharts (https://cran.r-project.org/web/packages/leaflet.minicharts/vignettes/introduction.html).

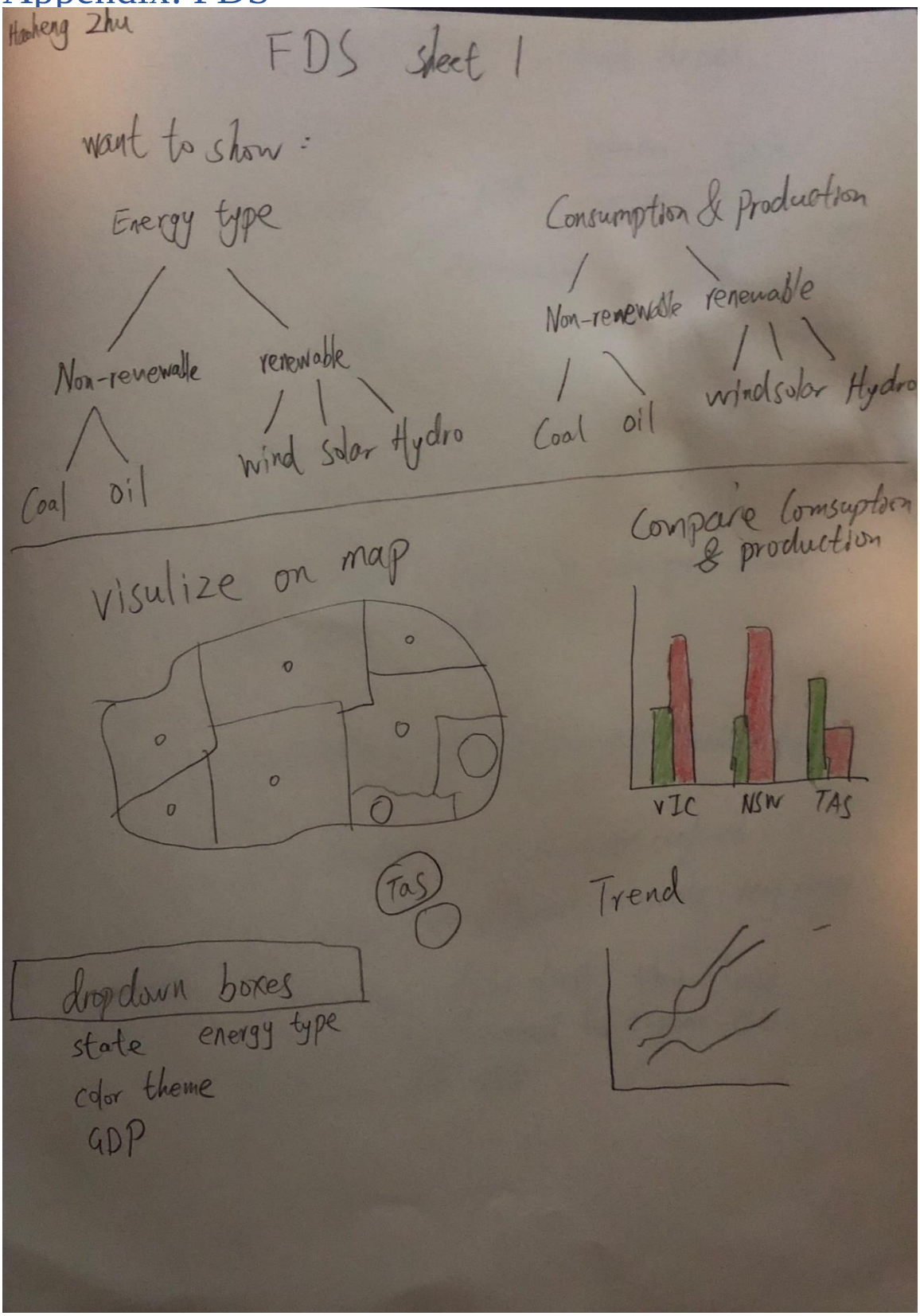
Minicharts for dynamic leaflet maps (https://github.com/rte-antares-rpackage/leaflet.minicharts).

Package 'readxl' (https://cran.r-project.org/web/packages/readxl/readxl.pdf).

Package 'VIM' (https://cran.r-project.org/web/packages/VIM/VIM.pdf).

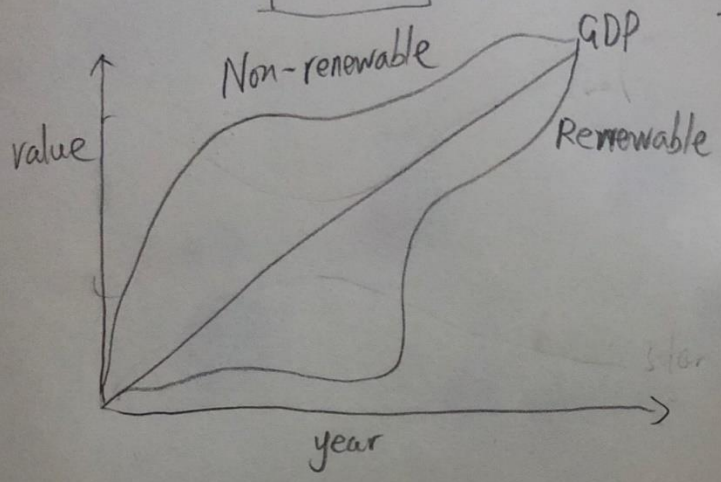
Visualize Missing Data with VIM Package (https://www.datacamp.com/community/tutorials/visualize-data-vim-package).

Appendix: FDS



FDS sheet 2 First design

selection option



selection: state

options:

Export / import
Energy type
Consumption
production
prediction

Pros: great for trend. selection can make graph cleaner.
Lots of freedom to user.

cons: cant visulize many category simultaneously. Doesn't refelct location.
Too many selections can make user confused.
If two selections are different in scale, very messy.

Operation: user selects state first, then choose energy or export/import to explore the correlation with GDP.

FDS sheet 3 second design

Cartogram Map

time

distortion is proportional to GDP.

Pie chart reflects Non-renewable and renewable.

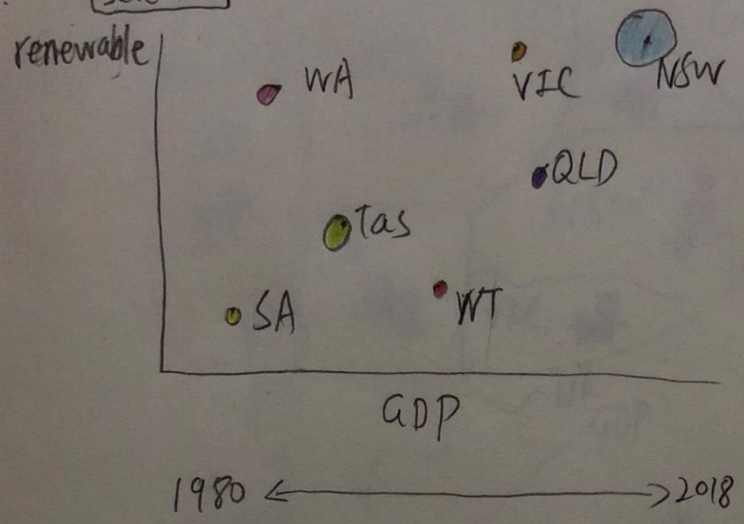
Pros: visulize loction, GDP, Energy type. Reveal correlation among GDP, Energy consumption

Cons: Doesn't reflect prediction. Doesn't reflect time series.

operation: select time period. 5 years as a period. take avg value of GDP/consumption.

FDS sheet 4 3rd design

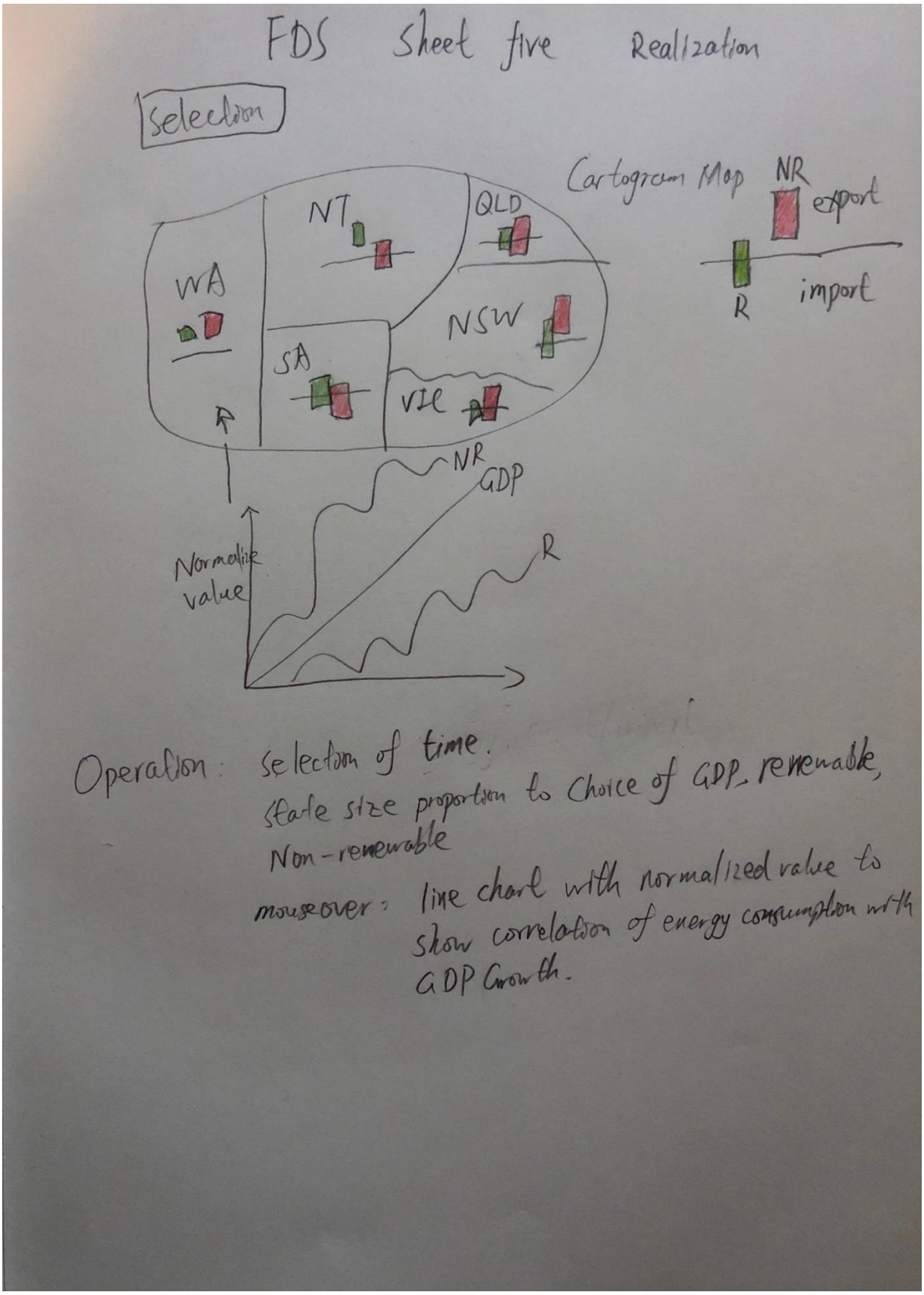
Motion Chart



Operation: selection of Energy type, Export/import on Y-axis.
bubble size reflects energy consumption/production.

Pros: Great for time series, visualize trend over time.
allow to compare multiple states simultaneously.

Cons: Doesn't show location, might be hard to understand the story at first glance, have to digest a while.
Can't show the proportion of renewable and non-renewable


FDS Sheet five Realization
Selection
Cartogram Map
NR
export
import
R
NT
QLD
WA
NSW
SA
VIC
NR
GDP
R
Normalize value
Operation: selection of time.
State size proportion to Choice of GDP, renewable, Non-renewable
mouseover: line chart with normalized value to show correlation of energy consumption with GDP Growth.