

Data analysis of people interested in Science in Australia

Yiheng Liu
Melbourne Australia
s3673551@student.rmit.edu.au

This report will mainly analyze two contents. Firstly, it will analyze people who are most interested in science, and analyze this gender, address, and age separately. In addition, this report will analyze people studied science in the school and whether they found jobs which are related to science latter. This report analyzes data through quantitative analysis and derives the impact of gender, age, and place of states on participants who are interested in science.

Keywords—data analyze, science, gender, address, age

I. INTRODUCTION

In the past decade, technology has developed rapidly, so people began to pay attention to science. Smartphone, computer, and other digital devices become a hot topic in daily life. In order to advance the development of science and surveyed the employment of science in Australia, the company organizes employees to conduct data analysis based on data provided by the Victorian Government Department of State Development. By studying the range of residence and age distribution and gender differences of people who most interested in science, the impact of science on gender, living place, and age can be derived. Analyzing the work situation of participants can summarize whether learning science is good for work. In the survey, a total of 3178 people participated in the survey; however, the data is old, some participants' data was lost. Therefore, only 3074 participants' data have been retained. All data were collected in DSDBI in 2011.

II. FINDINGS

A. Data cleaning

Because data are old, some data was lost. Therefore, in order to data integrity and reduce data errors, this report will take data cleaning. If the data have some problems, such as typing errors, repeated entries, format variations, missing data, fake data, for this problem, data cleaning is essential[1].

These data are taken from DSDBI. According to Figure 1, state data was missing. In this report, the data of state are very important, because living address must be analyzed. Due to losing state data, it is not possible to determine whether participants are living in Australia. Those data need to be cleaned. Therefore, those data need to remove.

ID	AGE	AGE BAND	AGE BAND 2	POSTCODE	State	SES
5	42	35-44yo	35+		#N/A	Unclassified
11	26	25-34yo	Under 35	1234	#N/A	Unclassified
45	35	35-44yo	35+		#N/A	Unclassified
58	47	45-54yo	35+		#N/A	Unclassified
59	49	45-54yo	35+		#N/A	Unclassified
70	60	55-64yo	35+		#N/A	Unclassified
78	29	25-34yo	Under 35		#N/A	Unclassified
139	58	55-64yo	35+		#N/A	Unclassified
178	41	35-44yo	35+		#N/A	Unclassified
182	62	55-64yo	35+		#N/A	Unclassified
203	31	25-34yo	Under 35		#N/A	Unclassified
226	31	25-34yo	Under 35		#N/A	Unclassified
295	27	25-34yo	Under 35		#N/A	Unclassified
302	46	45-54yo	35+		#N/A	Unclassified
359	63	55-64yo	35+		#N/A	Unclassified
364	28	25-34yo	Under 35	5140	#N/A	Unclassified
388	43	35-44yo	35+		#N/A	Unclassified

Figure 1 Example Of Invalid Data

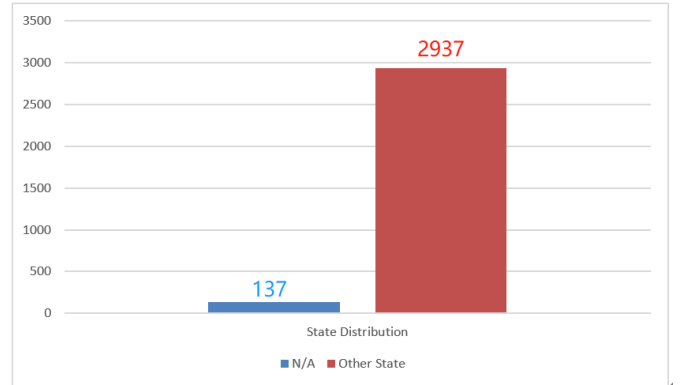


Figure 2 The number of useful data and missing data

Through the analysis of data, the number of useful data and missing data are obtained in Figure 2. There are 137 people loss state data. Therefore, this report will use other 2937 participants' data.

B. Statistics--Age

A total of 2937 people who participated in the survey and provided useful information. The table 1 shows that descriptive statistics of the participant age variable from the DSDBI data. The age of mean is 41.27, the minimum age is 10, maximum age is 88, the median age is 42, and the confidence interval is (40.715 to 41.842).

TABLE I. DESCRIPTIVE STATISTICS OF AGE

Descriptive Statistics	
Total	2937
Mean	41.27885598
Standard Error	0.287642581
Median	42
Mode	52
Standard Deviation	15.58852956
Sample Variance	243.0022537
Kurtosis	-0.855501774
Skewness	-0.018905469
Range	78
Minimum	10
Maximum	88
Sum	121236
Confidence Level (95%)	0.564001608
Lower bound	40.714854372
Upper bound	41.842857588

As shown in table 1, the average age of the people surveyed was 41.3 years old, which proves that most of the participants will be adults, and most people may have a job. From the data, the maximum age is 88 and the minimum age is 10. People of different ages are surveyed to enrich the diversity of data.

C. Data Analysis-Interest Level For Science

The survey divided the level of interest into five levels. There are “Very interested”, “Quite interested”, “Neither interested nor uninterested”, “Not very interested” and “Not interested at all” separately. “Very interested” is the highest level in this survey. It means that people at this level are the most interested in science.

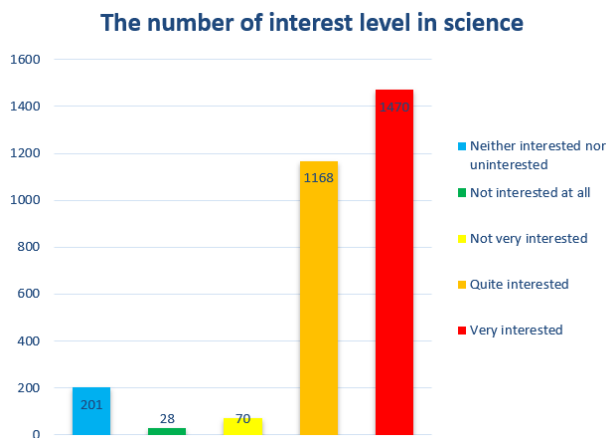


Figure 3 The number of interest level in science

Figure 3 shows that the most number is people who are very interested in science, which have 1470. It is interesting to note that people who choose “Quite interested” occupy nearly 40% of the total population in survey data, and they have 118. Therefore, it can be concluded that nearly 90% of the participants are interested in science. Therefore, science

D. Data Analysis-The group of participants who are most interested in science

According to figure 3, there are 1470 participants who are very interested in science. Therefore, it will use these data to analysis those people what age bands they are, and where states they lived, and what genders they are.

Age distribution of people who are very interested in science

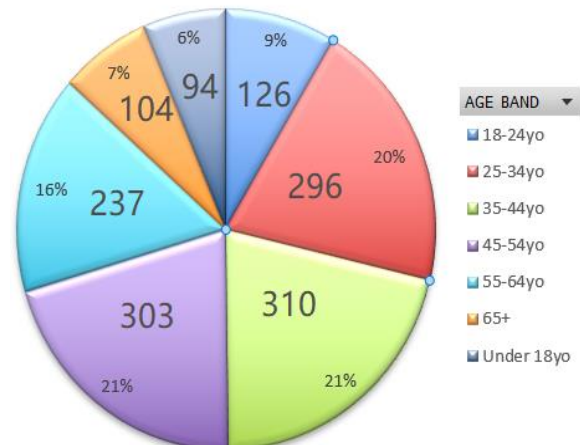


Figure 4 Age distribution of people who are very interested in science

According to figure 4, the number of participants in 25-34 years old and 35-44 years old and 45-54 years old are close, which are 296, 310, 303 respectively. They account for 60% of the total number. Participants' age in 25-54 years old has jobs based on science. Therefore, it may be a reason for people who were very interested in science. In addition, the number of participants whose age under 18 are the least, because they may study in primary school and secondary school, they just focused on their subject.

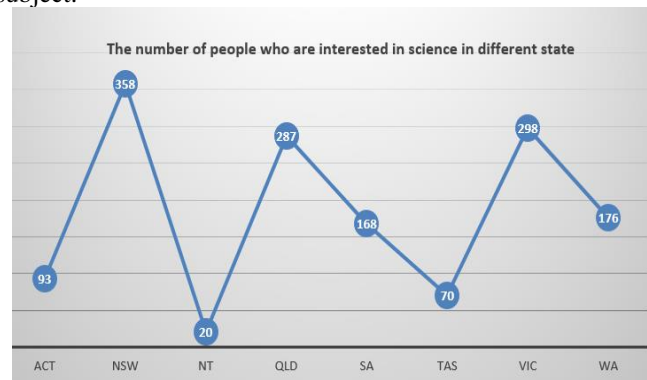


Figure 5 The number of people who are interested in science in a different state

It can be seen from figure 5 that participants who are interested in science living in NSW and QLD and VIC have more than 280. The largest number is in NSW, and their number is 358 because Sydney is Australia's largest city and the highest level of science. In addition, the number of participants living in QLD and VIC are similar, which are 287 and 298. It is interesting to note that NT only had 20 people who are very interested in science. The reason for this data, because NT is not a highly science area.

Compare the number of male and female

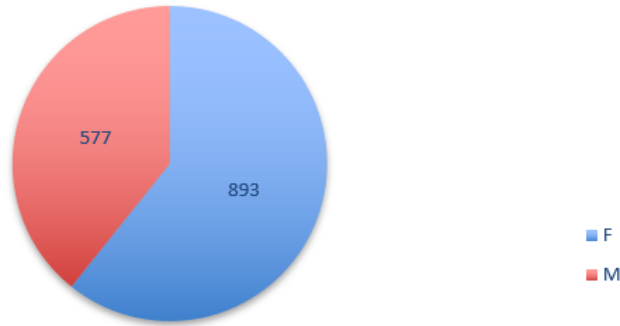


Figure 6 Compare the number of male and female

As can be seen from figure 6, there are 893 females who are very interested in science, males who have 577 are 316 less than female.

Each age band in each state

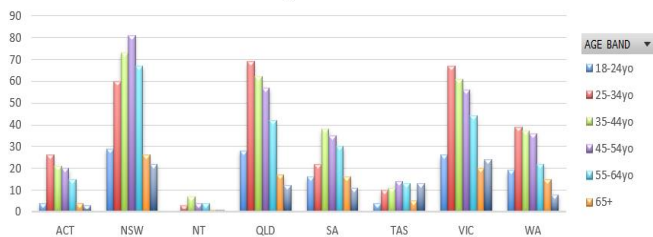


Figure 7 Each age band in each state

According to figure 7, participants living in NSW and aged 45-54 are the most, and they have 81 people. In addition, participants living in NSW and aged 35-44 have 73, which is the second largest number in the data. Furthermore, ACT, QLD, VIC, WA have an interesting commonality, which are the most people aged in 25-34. Finally, regardless of the state, the number of people aged band in 25-54 is the highest. Therefore, it can be concluded that in any region between the ages band of 25-54, they have the largest number of people who are interested in science.

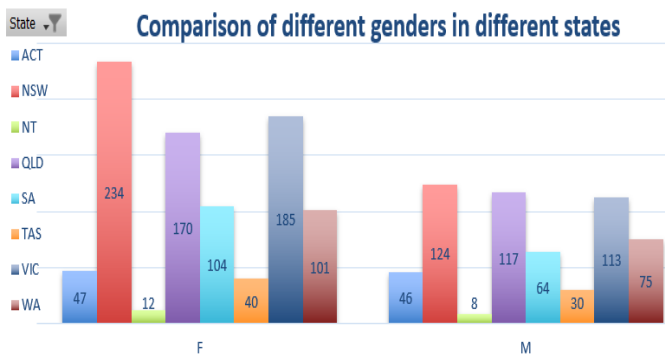


Figure 8 Comparison of different genders in different states

Figure 8 shows the gender and state connections in these data. Firstly, in NSW, no matter female and male, people

who are very interested in science are the most. In addition, no matter in which state, females who are very interested in science are more than male. It is interesting to note that males and females who are very interested in science living in NT are the least.

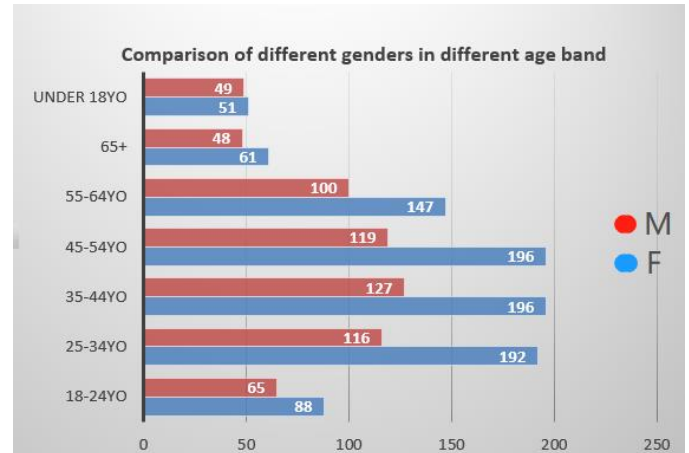


Figure 9 Comparison of different genders in the different age band

According to figure 9, the gender and age band connections in these data. In each age band, the number of females who are interested in science are more than males. The number of male and female whose age band in 35-44 and 45-54 are more than males and females of another age band.

From the above data, it can be concluded that the largest number of participants who are interested in science are females whose age band in 35-44, and they lived in NSW.

E. Data Analysis-work content

Excluding useless data, this data analysis will use the remaining available data which have 2937 for analysis. This data analysis will analyze people who had studied science and whether find work related to science in their career.

Education background for science

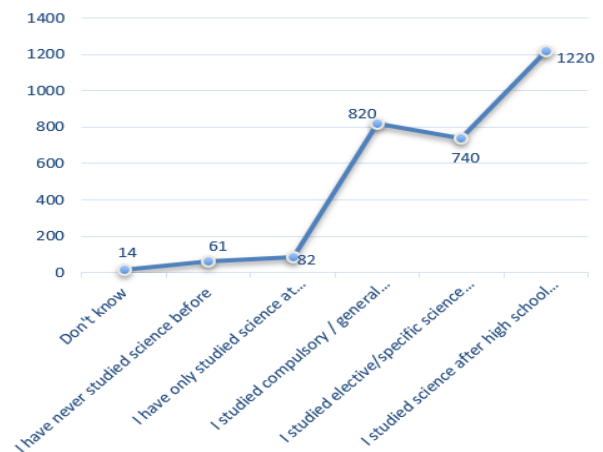


Figure 10 Education background for science

Figure 10 shows that six options in the questionnaire. There are "Don't know", "I have never studied science

before”, “I have only studied science at primary school”, “I studied compulsory/ general science subjects at school”, “I studied elective/ specific science subjects at school” and “I studied science after high school at TAFE or Uni” respectively. The number of participants who chose them was 14, 61, 82, 820, 740, 1220.

As can be seen from figure 10, they have a total of 2862 people who have studied science in different degrees. More than 1200 people started studying science after high school, which occupy 41% of the total participant. Therefore, this shows that in the Australian education system, most people would learn science in different levels of education in Australia.

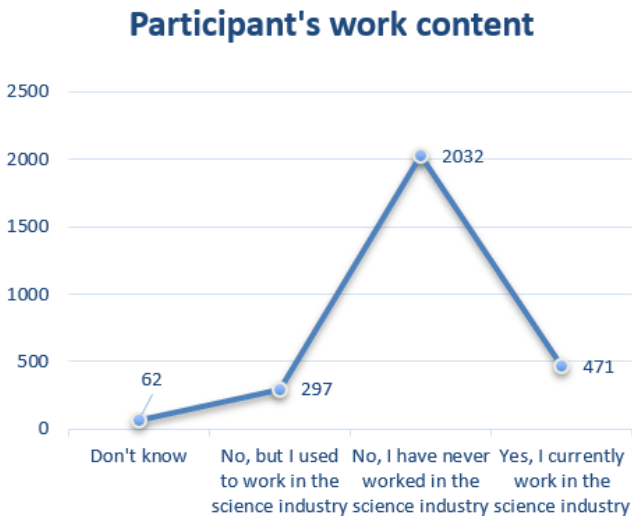


Figure 11 Participant's work content

In this case, it can be used 2862 people who studied science in different education degrees, because it will be analyzed by people who have studied science before, and use the data of these people to analyze whether their work base on science. According to the questionnaire, participants can choose 4 answers, which are “Don’t know”, “No, but I used to work in the science industry”, “No, I have never worked in the science industry”, and “Yes, I currently work in the science industry”

According to figure 11, there are 2032 people found a job without science. It is interesting to note that only 471 people found a job in the science industry at that time. Finally, there are 62 people who do not know, because they may study in school.

III. OBSERVATIONS

- 1) More females are very interested in science than males.
- 2) The number of people who like science in NSW is more than other states.
- 3) Most people who are very interested in science most are between 25-54 years old.

- 4) Most people have studied science in different education degrees.
- 5) More than 75% of people whose job are not related to science.

IV. CONCLUSIONS

To conclude, a lot of data will be generated when the survey is conducted. Before data analysis, data cleaning must be performed to eliminate unnecessary or invalid data. These data were from 2011, so some data were lost, and the questionnaire cannot be found from the government online, which may lead to errors in data analysis. Therefore, in the future data analysis, it should have all the information, including data, questionnaires.

REFERENCES

- [1] R. Almeida, P. Oliveira, L. Braga, and J. Barroso, “Ontologies for Reusing Data Cleaning Knowledge” 2012 IEEE Sixth International Conference on Semantic Computing, Sep 2012