TOPIC 9
DATA INTERPRETATION

Don't be a novelist – be a statistician…. much more scope for the imagination.

Huffi

Introduction

The role of a statistician is not only to produce statistics but also to explain them. The interpretation of statistical data is both an art and a science. We need to correctly identify the main facts presented in tables and graphs, and understand the meaning behind them. This requires the application of both statistical skills and real-world knowledge and experience. In this topic we will learn how to read statistical tables and graphs, how to write up the analysis and how to interpret the results.

Reading Statistical Tables and Graphs

The first thing we look at in any table is the title to find out what information the table contains. The heading should tell us what statistical units are being presented (e.g. persons, households, dollars, etc), the time period (e.g as at a date, quarter, year, etc), and how the data is classified (i.e. the variables). Also, if the table has a headnote we should make note of what units (the observations) the table is reporting (e.g. '000 persons). For example, the title below indicates that we are looking at a table of employed persons broken down by two variables: type of employment and state, for the Federated States of Micronesia for the year 1997.

Example:

Employed Persons by Type of Employment and State, FSM 1997

Having noted what information is contained in the table, we now need to look at the caption (column headings) and stub (row headings) to find out what and how many particular variable values are presented. The example below shows there are four types of employment and four states.

Example:

Type of Employment	Yap	Chuuk	Pohnpei	Kosrae	Total
Private Wage and Salary					
Government					
Self-Employed					
Unpaid Family Work					

The next stage is to examine the body of the table; in particular we want to see if we can identify any trends or pattern in the numbers and also note any cells that seem peculiar or different from the others. We can do

this kind of very basic analysis on the table as it stands, but it may be a little difficult to really see what is happening. We should take note of how totals are calculated and read any footnotes specified.

Example:

Type of Employment	Yap	Chuuk	Pohnpei	Kosrae	Total
Private Wage & Salary	1,401	4,678	5,191	558	11,828
Government	3,484	9,271	5,307	1,035	19,097
Self-Employed	108	555	511	-	1,173
Unpaid Family Work	-	43	23	-	66
Total Employed	4,992	14,547	11,032	1,593	32,164

Source: 1997 FSM Labor Force Survey Note: Totals may not add due to rounding.

If we wish to use the table to make comparisons and draw conclusions then we could rewrite the table by reordering the rows and columns so that the largest values come first. Note however that this is not a good idea for ordinal or interval variables since they have a logical order, or for many nominal variables where the order is pre-defined by convention. For example, geographical variables are often ordered north to south or west to east.

Example:

Employed Persons by Type of Employment and State, FSM 1997

Type of Employment	Yap	Chuuk	Pohnpei	Kosrae	Total
Government	**3,484**	**9,271**	**5,307**	**1,035**	**19,097**
Private Wage & Salary	1,401	4,678	5,191	558	11,828
Self-Employed	108	555	511	-	1,173
Unpaid Family Work	-	43	23	-	66
Total Employed	4,992	14,547	11,032	1,593	32,164

Source: 1997 FSM Labor Force Survey Note: Totals may not add due to rounding.

We can see from the above table, which has been sorted by row, that the Government employs the largest number of workers, with most employed in Chuuk State. But the proportion of Government employees to the total employed labour force varies between states. It would be easier to interpret these results if the numbers in the table were presented as percentage of the total for each state. The tables below shows that the largest percentage of Government employees are employed in Yap State.

Example:

Employed Persons by Type of Employment and State, FSM 1997
(Percentage of State Total)

Type of Employment	Yap	Chuuk	Pohnpei	Kosrae	Total
Government	**69.8**	**63.7**	**48.1**	**65.0**	**59.4**
Private Wage & Salary	28.1	32.2	47.1	35.0	36.8
Self-Employed	2.2	3.8	4.6	-	3.6
Unpaid Family Work	-	0.3	0.2	-	0.2
Total Employed	100.0	100.0	100.0	100.0	100.0

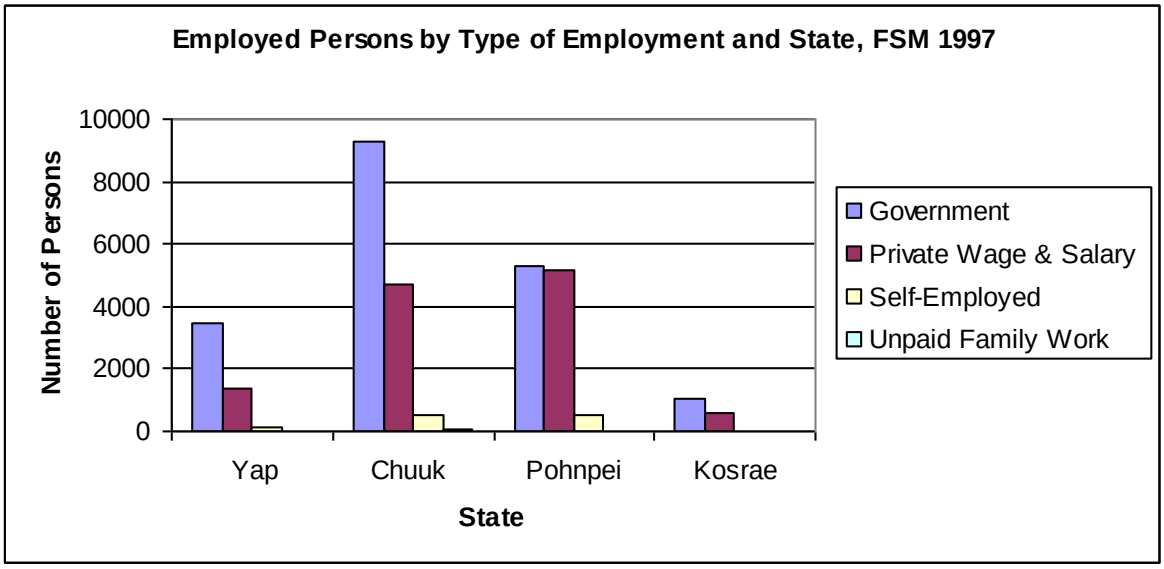
Source: 1997 FSM Labor Force Survey Note: Totals may not add due to rounding.

Graphs are more revealing than statistical tables as the principal function of a graph is to present data in an unambiguous and illuminating form. Graphs emphasise to the reader the main statistical message. As with

tables you should make note of the titles, labels and annotations presented in a graph. You should also note that whether the graph uncovers patterns and relationships in the data that may need to be commented on in the analysis.

For example, the graph below emphasises the largest number of employees in each state are Government workers, and shows the largest number are employed in Chuuk State:

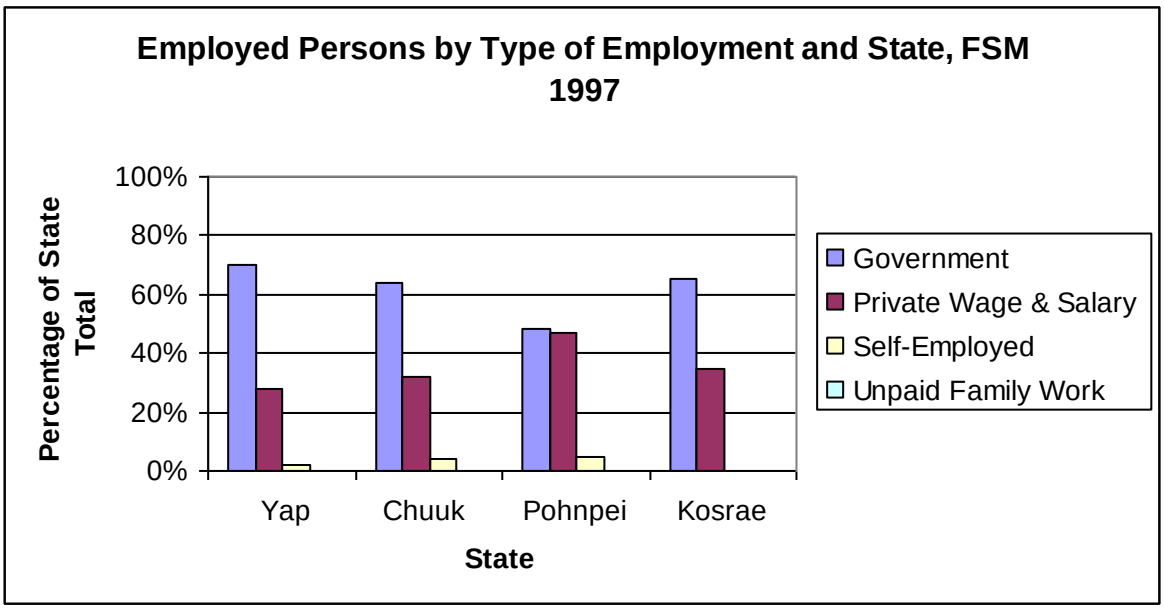
Example:



Source: 1997 FSM Labor Force Survey

If we wanted to comment on the proportion of government workers in each state then a percentage graph would be more appropriate. As shown below, the highest percentage of government workers were in Yap State with the fewest in Pohnpei State. We should also note that Pohnpei State has a similar proportion of government employees to private wage and salary workers.

Example:



Source: 1997 FSM Labor Force Survey

TIP

Graphs and tables should illustrate the key messages you wish to convey to the reader. Too much information in a table or graph will confuse the reader!. Each table and graph should make a point, much like a paragraph.

Writing Up The Analysis

Now, we are ready to describe some of the results in our table or graph. A key approach in writing your report should be to describe and explain the technical terms used in your analysis so that the report can be easily understood by the reader.

It is common practice to refer to tables or graphs in text by their table or figure number. For example:

- As shown in figure 1, ……..
- It can be seen from table 2 that…..
- Figure 2 refers to …..

These references imply that the reader has yet to come across the specific tables or figures, and so the text should come before the presentation of the tables and graphs. It is also acceptable to reference the table or figure number in parenthesis so that it does not obscure the main point of the message (e.g. see figure 1).

It is generally not useful to restate the numbers presented in a table or graph in text as the reader will have already have seen the table. However in cases where the size of the values is small (e.g. less than 50) then the frequency count should only be reported.

Generally, it is good practice to explain the size of the values in terms of percentages as this is a concept the audience easily understands. Remember to describe both the largest and lowest values, and also the most common values. Also, mention any unusual values that don't fit the expected pattern.

Example:

> Figure 11 looks at the educational pattern by attainment. The figure shows that females tended to be less educated than their male counterparts. Females are more likely to have below high school attainment (68 percent) than males (38 percent).
>
> *Johnson BS A Summary Analysis of Selected Indicies of the Prelimary Results of the Kosrae 2000 census.*

In some tables it may be possible to aggregate categories in order to make the explanation more clear or meaningful. This is particular useful in tables which have ordinal or numeric variables, such as attitudinal scales or age groups.

Example:

> When main daily activity is examined in terms of age group, it is clear that most of the 10-19 age group are full-time students, with a small proportion of this age group involved in economic activities. It is interesting to note the significant number of people aged above 50, especially women, who indicated that domestic duties was their main activity. Usually this group is classified as 'not applicable' because they are retired or do not spend most of their time doing domestic duties.
>
> *Unpaid Household Activities Report. Household Income and Expenditure Survey. Department of Statistics, Government of Samoa 1997*

It may be useful to describe a normal distribution in terms of an average and standard deviation. These statistics are often presented in the text, rather than in a table, as their meaning do require some explanation. For non-normal data you may want to describe the shape of the distribution and its skewness.

Example:

> Both populations produced were of size 2000 businesses. The population mean for the normally distributed population with a standard deviation of 3000 was 12416 while for the positively skewed population it was 13780..…
>
> *Impact of non-response bias in business surveys, Australian Bureau of Statistics, 1999*

While your reporting of the results should be based on the facts presented, you should refer to possible reasons for any unusual findings. For example, the reasons for a sudden change in a time series may be known and should be commented on in the analysis.

Example:

> As figure 5.5 shows, in 1986 over half the Pacific population (54 percent) had no formal educational qualifications. By 2001 this had fallen to 36 percent…Much of the trend has been due to the increasing attainment of school qualifications.
>
> *Pacific Progress: A report on the economic status of Pacific Peoples in New Zealand. Statistics New Zealand 2002.*

Also, if other analyses have shown that a statistical event or trend has occurred as a result of a particular relationship with an independent variable or other dependent variables then this should be reported in support of your analysis.

Example:

> A simple way to investigate the association between causes of death is to calculate the percentage of deaths involving a particular cause which also involve another specified cause….An area where this form of analysis might be particularly useful is cross-classifying instances of suicide by mental health problems, as set out in table 2. From this table it is clear, for example, that the proportion of suicides where mood affective disorders are also mentioned is substantially higher for females than males.
>
> *Multiple causes of death analysis, Australian Bureau of Statistics, 2003*

In some cases it may not be possible to find an explanation for a particular statistical result or trend. In these cases, you may want to mention there is a need for more data to be collected or further analysis to be performed. Such data or analysis might be able to provide a realistic explanation or at least confirm a suggested one.

Example:

> The household, based on HIES results, is still an extremely important domain in Samoa. Additional research is required to quantify and assess the relative importance of the other domains in relation to the household. This could be done reliably through a time use survey.
>
> *Unpaid Household Activities Report. Household Income and Expenditure Survey. Department of Statistics, Government of Samoa 1997*

TIP

It is a good idea to begin writing descriptions of the analysis and noting supporting evidence as you complete each section of the data analysis. This will assist you when you come to write the draft report.

Interpreting The Results

The interpretation of statistical data requires experience and a common sense approach. To a large extent the interpretation of results will be influenced by the needs of the audience, the objectives of the research and the quality of the data presented.

The interpretation of your findings are presented in the discussion section of the report. You must interpret your results for the readers so that they can understand the meaning of your findings. The two most important questions to answer are:

- ? What do my results mean?
- ? What are their implications?

The Interpretation

The results of the analysis should be interpreted for the reader in terms of the objectives set out in the introduction. The key questions to be answered are:

- ? What are the overall findings from the analysis?
- ? How reliable or valid are the findings?
- ? What explanation can be provided for the findings?

A summary of the results should be provided in the discussion section. These may be presented in dot point fashion as shown below. The summary should refer to the objectives of the study and show how the results are relevant to meeting these objectives. It is important to ensure that the explanations you present are supported by the statistical data.

Example:

The statistics on the businesses in the formal sector show:

- The concentration of formal sector businesses and employment in the service industries, rather than the primary or goods producing industries.
- The low number of formal sector businesses in the agricultural and forestry industry. This reflects the nature of the Vanuatu economy with a significant portion of the population involved in household level agricultural activities outside the formal sector.
- The significance of the government sector in terms of formal sector employment numbers.
- The high number of privately owned businesses in the small size category of 1-9 employees.
- The high concentration of privately owned businesses in the two urban areas of Port Vila and Luganville.

Vanuatu 2000: Labour Market Survey Report. Vanuatu Statistic Offfice, 2000

The statistical relationship between variables should be commented on where cross-tabulations have been produced. However care needs to be taken to ensure that inferences about the relationships are supported by the data. In particular the nature of the relationship between independent and dependent variables need to be considered – that is, how strong and in what direction are the associations between the variables. Note that in social surveys it is usual to refer to associations between variables, rather than to casual relationships.

Consideration of data quality is especially important in interpreting results of both surveys and administrative data sources. Unusual results or unexpected events that influence the data should be explained to readers to help them in the interpretation of the findings. In the discussion of the results it is good practice to provide a summary of the issues that affect the validity and reliability of the results. Note that comprehensive measures of data quality should be provided in the appendix (refer to the following information box for detailed discussion of the assessment of data quality).

Example:

> With respect to the Solomon Islands, developments already superseded part of the information captured in the 1999 census. The country experienced two major upheavals connected to the current social unrest…This especially affected the situation with regard to the geographical distribution of the population and economic activities.
>
> *Solomon Islands 1999 Population and Housing Census: Main Results*

The Implications

You should also discuss the relevance of your findings to the wider research in the area of study. The key questions that should be answered are:

- **?** Does it support the findings of previous studies?
- **?** What recommendations can be made from the study?
- **?** What further analysis could be done?

In the background to your study you should have referred to other research in the field. It is important to show how your study has contributed to the body of research by either confirming or negating previous findings. You should also state any recommendations that come from your study. However it is very important to ensure that the recommendations you present are supported by the statistical data. In some cases you will need to indicate that your study is not conclusive and that further work needs to be done.

Example:

> As they continue to make up an increasing proportion of the population, and of the workforce in particular, the position of Pacific peoples will assume increasing social and economic significance. The implications of this population growth are of major importance not just for the Pacific communities but for New Zealand as a whole. If the positive trends identified in this report are built on in the future and the current disparities are addressed, Pacific people will be in a better position to take advantage of the opportunities and challenges ahead and New Zealand as a whole will benefit.
>
> *Pacific Progress: A report on the economic status of Pacific Peoples in New Zealand. Statistics New Zealand 2002.*

In the conclusion, you may wish to comment on the availability of relevant statistical data within your country and within the region. You could point out what gaps exist in the current statistical collections and what improvements could be made to provide the necessary data in the area of study. However, it is important to research the area thoroughly and consult with the various agencies responsible for producing statistical information to ensure that your comments are valid.

Example:

It has been argued that government statistics do not adequately reflect issues which concern women. This is a problem because governments base their decisions on research data, and good public policy is based on good statistics. In many Pacific Island countries and territories there is a lack of statistics on availability and use of credit by men and women, allocation and control of resources in the household and division of family and household responsibilities. Governments need data in these and other relevant areas so that they can develop policies and programmes that address the concerns raised in the Pacific Platform for Action.

Getting Started in Statistics: A handbook for Pacific Women leaders. Secretariat of the Pacific Community 1998

Caution

Try to avoid making recommendations that are not directly related to your area of study and that are not supported by statistical evidence. In many statistical reports recommendations are not made, but rather the implications of the findings for decision makers are stated.

.

Information Box: Assessment of Data Quality

In general you should aim to identify the most important sources of error and provide quantitative measures where possible or qualitative descriptions otherwise. The result should be a balanced discussion which addresses itself to specific sources of error or bias and is therefore informative to readers.

The following aspects of data quality should be provided where applicable (based on the protocols for official statistics, Statistics New Zealand):

1. ***Coverage and scope:*** the quality of the survey frame for surveys or censuses, or source files for administrative data should be discussed (including gaps, duplications and definitional problems).

2. ***Sampling error:*** if the survey is based on a random sample then estimates of the sampling error of tabulated data based on the sample should be provided, together with an explanation of how these standard error figures should be used to interpret the data.

3. ***Response rates:*** the percentage of the target sample or population from which responses or usable data were obtained should be provided. Any known differences in the characteristics of respondents and non-respondents should also be described.

4. ***Comparability over time:*** it may be appropriate to discuss comparability with the results of the same activity for a previous reference period, especially if there has been a change in methodology, concepts or definitions.

5. ***Benchmarking and revisions:*** the effects of benchmarking or revisions on comparability over time should be described. Guidance on the possible impact of future benchmarking should be given based on past experience.

6. ***Comparability with other data sources:*** if similar data from other sources exist they should be identified. Where appropriate, a reconciliation should be attempted describing how the data sets differ and the reasons for these differences.

7. ***Response bias***: an assessment of the effect of response bias on the statistics should be provided if possible. Evidence of response bias problems stemming from respondent misunderstanding, questionnaire problems, or other sources should be provided if available.

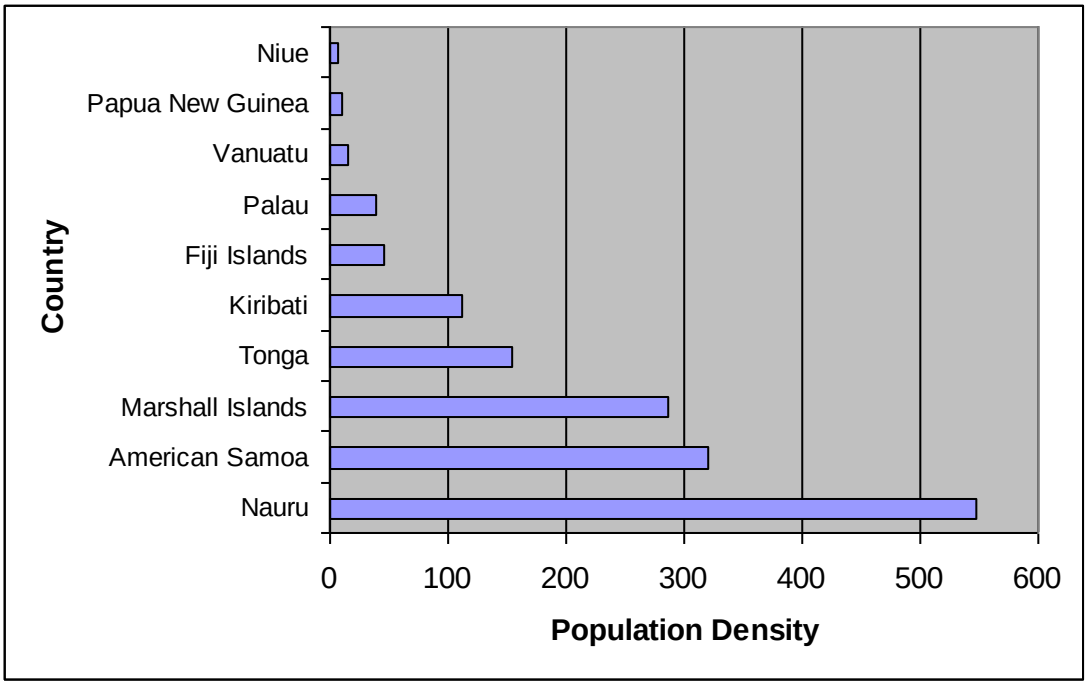
8. ***Edit and imputation*** : the effect of editing and imputation on the quality of data should be assessed and described.

9. ***Other error sources:*** if there are particular sources of error or unforeseen events which are relevant to the series or occasion, these should be described.

... Exercises ...

Go through the following tables and graphs and write a brief paragraph describing what they are showing. Note any unusual patterns in the data.

(a) Population Density for selected countries in the Pacific, 2000



Brief description :

.....

.....

.....

.....

.....

.....

.....

.....

.....

b) Total number of victims from the last four Fiscal years, 1997–2000, American Samoa

Type of crime	1997	1998	1999	2000	Total
Child physical abuse	605	600	188	975	2368
Child sexual abuse	445	390	77	565	1477
DUI/DWI	0	0	58	25	83
Domestic Violence	87	185	249	80	601
Adult Sexual Assault	21	10	42	40	113
Elder Abuse	23	5	185	10	223
Adults Molested as Children	11	8	0	15	34
Survivors of Homicide Victims	0	50	14	10	74
Robbery	0	50	10	20	80
Assualt	12	5	334	35	386
Child Neglect	0	0	26	3	29
Suicide Attempt	0	0	16	2	18
TOTAL	1204	1303	1199	1780	5486

Brief description :

.....

.....

.....

.....

.....

.....

.....

.....

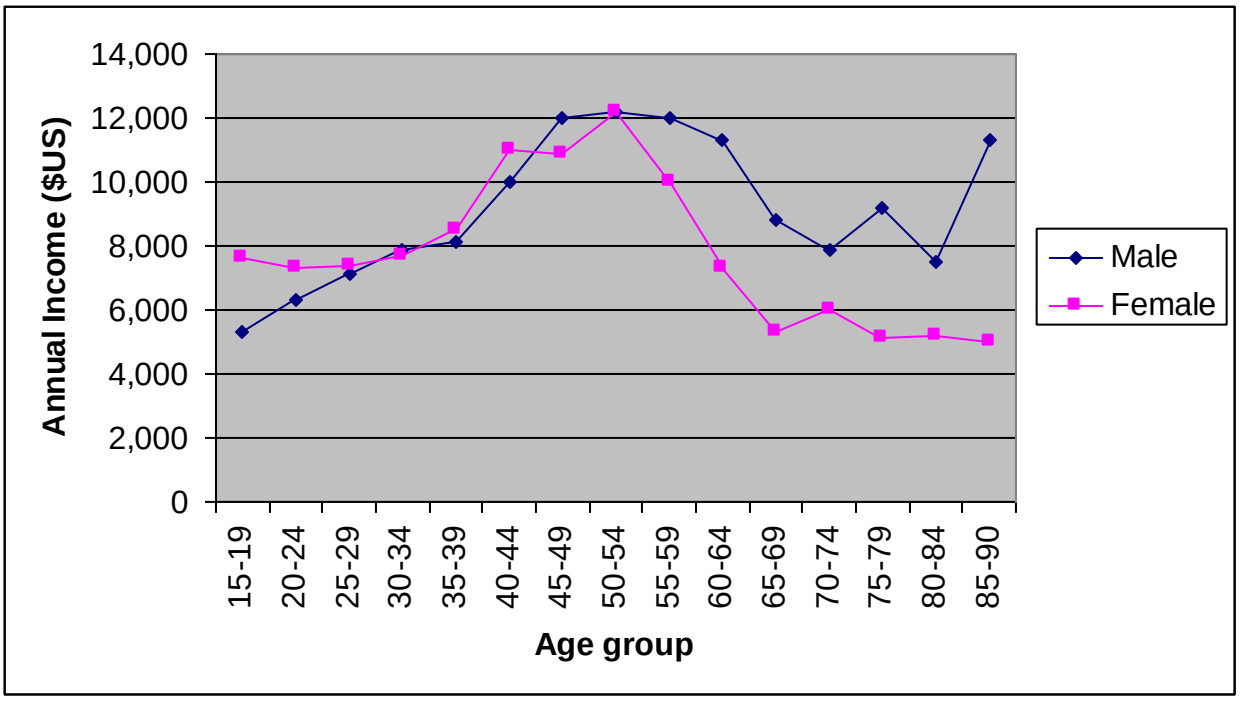
.....

.....

.....

.....

c) Average Annual Income of people with Secondary School Educational Attainment, by age group and gender



Brief Description :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....