**ASSESSMENT COVER SHEET**

|  |  |  |  |
| --- | --- | --- | --- |
| Unit Name and Code: | **FIT5141 Advance Topic of Information Technology** | | |
| Campus: | **Caulfield** | | |
| Assignment Title: | **Deployment of Dataset on MongoDB and R based simple analysis of Data** | | |
| Name of Lecturer: | **Ariel Liebman** | | |
| Name of Tutor: | **Muhammad Osama** | | |
| Tutorial Day and Time: | **Tuesday 10:00am - 12:00am** | | |
| Phone Number: | **0490778801** | | |
| Email Address: | **sfen0004@student.monash.edu** | | |
| Has any part of this assignment been previously submitted as part of another unit/course?  Yes  No | | | |
| Due Date: | **Sep 13th 2017** | Date Submitted: | Sep 13th 2017 |
| All work must be submitted by the due date. If an extension of work is granted this must be specified with the signature of the lecturer/tutor.  Extension granted until (date)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Signature of lecturer/tutor \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Please note that it is your responsibility to retain copies of your assessments. | | | |
| ***Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations***  **Plagiarism**: Plagiarism means taking and using another person’s ideas or manner of expressing them and passing them off as one’s own. For example, by failing to give appropriate acknowledgement. The material used can be from any source (staff, students or the internet, published and unpublished works).  **Collusion**: Collusion means unauthorised collaboration with another person on assessable written, oral or practical work and includes paying another person to complete all or part of the work.  Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or delegate, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing. | | | |
| **Student Statement:**   * I have read the university’s Student Academic Integrity [Policy](http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-academic-integrity-policy.html) and [Procedures](http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-academic-integrity-managing-plagiarism-collusion-procedures.html). * I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations http://adm.monash.edu/legal/legislation/statutes * have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied. * No part of this assignment has been previously submitted as part of another unit/course. * I acknowledge and agree that the assessor of this assignment may for the purposes of assessment, reproduce the assignment and:  1. provide to another member of faculty and any external marker; and/or 2. submit it to a text matching software; and/or 3. submit it to a text matching software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.  * I certify that I have not plagiarised the work of others or participated in unauthorised collaboration when preparing this assignment.   Signature ..........Siyang Feng..................................... Date……… Sep 13th 2017……………………  \* delete (iii) if not applicable | | | |

|  |  |
| --- | --- |
| **Student ID number** | 28246993 |
| Given Name | Siyang |
| **Family name** | Feng |

*The information on this form is collected for the primary purpose of assessing your assignment* and ensuring the academic integrity requirements of the University are met*. Other purposes of collection include recording your plagiarism and collusion declaration, attending to course and administrative matters and statistical analyses. If you choose not to complete all the questions on this form it may not be possible for Monash University to assess your assignment. You have a right to access personal information that Monash University holds about you, subject to any exceptions in relevant legislation. If you wish to seek access to your personal information or inquire about the handling of your personal information, please contact the University Privacy Officer:* [*privacyofficer@adm.monash.edu.au*](mailto:privacyofficer@adm.monash.edu.au)

FIT5141 Advance Topic of Information Technology

*Assignment 2*

*Siyang Feng*

*28246993*

**Catalogue**

[Part 1. Implementation 5](#_Toc493079093)

[Task 1. Building MongoDB server 5](#_Toc493079094)

[Task 2. Make connection with MongoDB and R 6](#_Toc493079095)

[Task 3. Make connection with MongoDB and Tableau 7](#_Toc493079096)

[Part 2. Statistics and Charts 8](#_Toc493079097)

[*Figure 1 Command of active MongoDB 5*](#_Toc493078999)

[*Figure 2 Open the Mongo database 5*](#_Toc493079000)

[*Figure 3 Import csv data into Mongo database (for loop and regular expression) 5*](#_Toc493079001)

[*Figure 4 Problem of importing data 6*](#_Toc493079002)

[*Figure 5 Import collection from MongoDB into R environment 6*](#_Toc493079003)

[*Figure 6 Environment collections in R 6*](#_Toc493079004)

[*Figure 7 Generate schema definition file 7*](#_Toc493079005)

[*Figure 8 Start mongosqld 7*](#_Toc493079006)

[*Figure 9 Connect Tableau and MongoDB database 7*](#_Toc493079007)

[*Figure 10 Pedestrian traffic (Melbourne Central) on weekends from 2011 to 2014 8*](#_Toc493079008)

[*Figure 11 Mean value in a day in 2014 8*](#_Toc493079009)

[*Figure 12 Pedestrian density in a day 9*](#_Toc493079010)

[*Figure 13 Pedestrian density in weekends 9*](#_Toc493079011)

[*Figure 14 Pedestrian density of four stations in 2014 10*](#_Toc493079012)

**Part 1. Implementation**

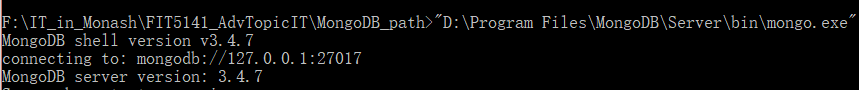
**Task 1. Building MongoDB server**

MongoDB installation package was downloaded into windows 10 system. Install the package and then open the CMD in admin authority. Make a new directory named data and in the directory, make another new directory named db. it is the database for mongo server. Then, connect the mongo server using the exe file, mongod.exe, with the “/data/db” directory.

屏幕剪辑

**Figure 1 Command of active MongoDB**

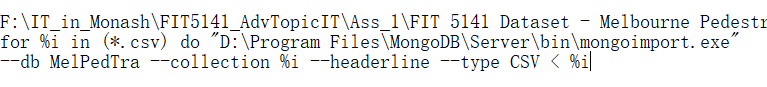
This command is used to connect MongoDB and the path directory. I have made it as a bat file. Each time starting Mongo, there is no need to input the start command.

When the server connection was done. Next step is open the mongo database. 

**Figure 2 Open the Mongo database**

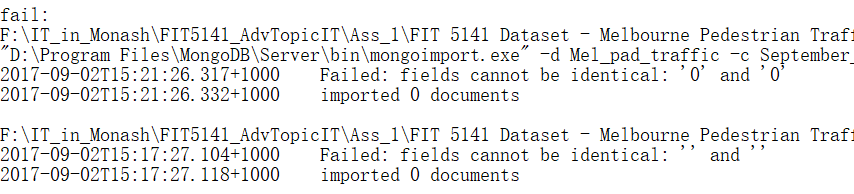
Open another CMD and active the mongo.exe, there will show the connecting result. From the figure above, the host IP is 127.0.0.1 with the port 27017 which will be used in connecting Tableau later.

Then, the csv noSQL collections will be imported into MongoDB using mongoimport.exe command. However, because I have thirty collections in my database. It will spend a lot of time on importing collection. Thus, I use the command, for loop and regular expression, in CMD to complete.



**Figure 3 Import csv data into Mongo database (for loop and regular expression)**

During the collection importing, I have a problem that some collections cannot import into server.



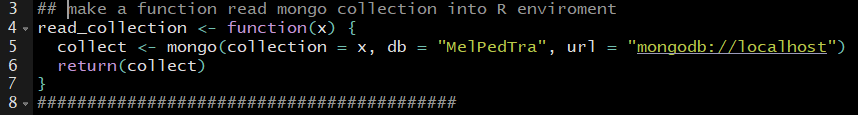
**Figure 4 Problem of importing data**

It is almost because that the encoding of collections has some problem. I solve this problem by opening the CSV file with excel and restore it to CSV format.

Some people have the authority problem and cd into bin directory and run the command will avoid this problem.

**Task 2. Make connection with MongoDB and R**

Keep the MongoDB server open and open Rstudio and then, input command, install.packages("mongolite"), which is the newest package to connect R with MongoDB.



**Figure 5 Import collection from MongoDB into R environment**

Figure 5 is the self-made function to import collections from mongoDB into R environment. In my server, all my collections are in the database named MelPedTra. Each collection in MongoBD should be each an environment in R.



**Figure 6 Environment collections in R**

The imported data was in the JSON data structure. Command, find(‘{}’), will be used to search data from a JSON collection.

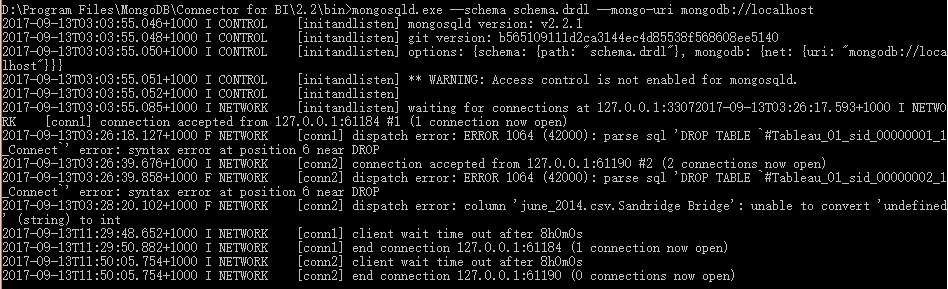
**Task 3. Make connection with MongoDB and Tableau**

The last part is to implement the connection between. To implement this task, MongoDB BI connector should be download for connecting. Install the MongoDB BI connector and run it.

屏幕剪辑

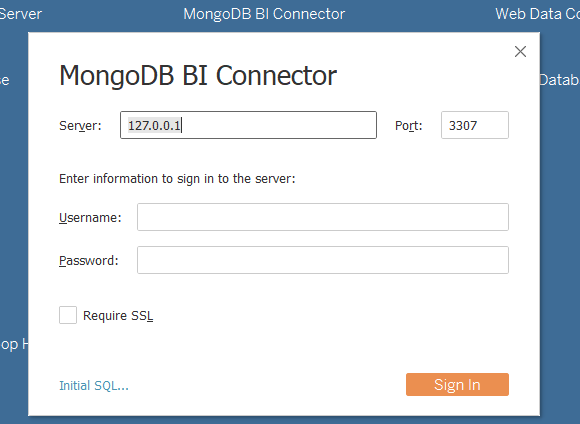
**Figure 7 Generate schema definition file**

Figure 7 indicates the command to generate schema definition file using mongodrdl to connect with database MelPedTra.



**Figure 8 Start mongosqld**

Figure 8 shows the start of mongosqld server. After that, open Tableau and in the “To a Server” bar, select more to find Mongo BI connector.

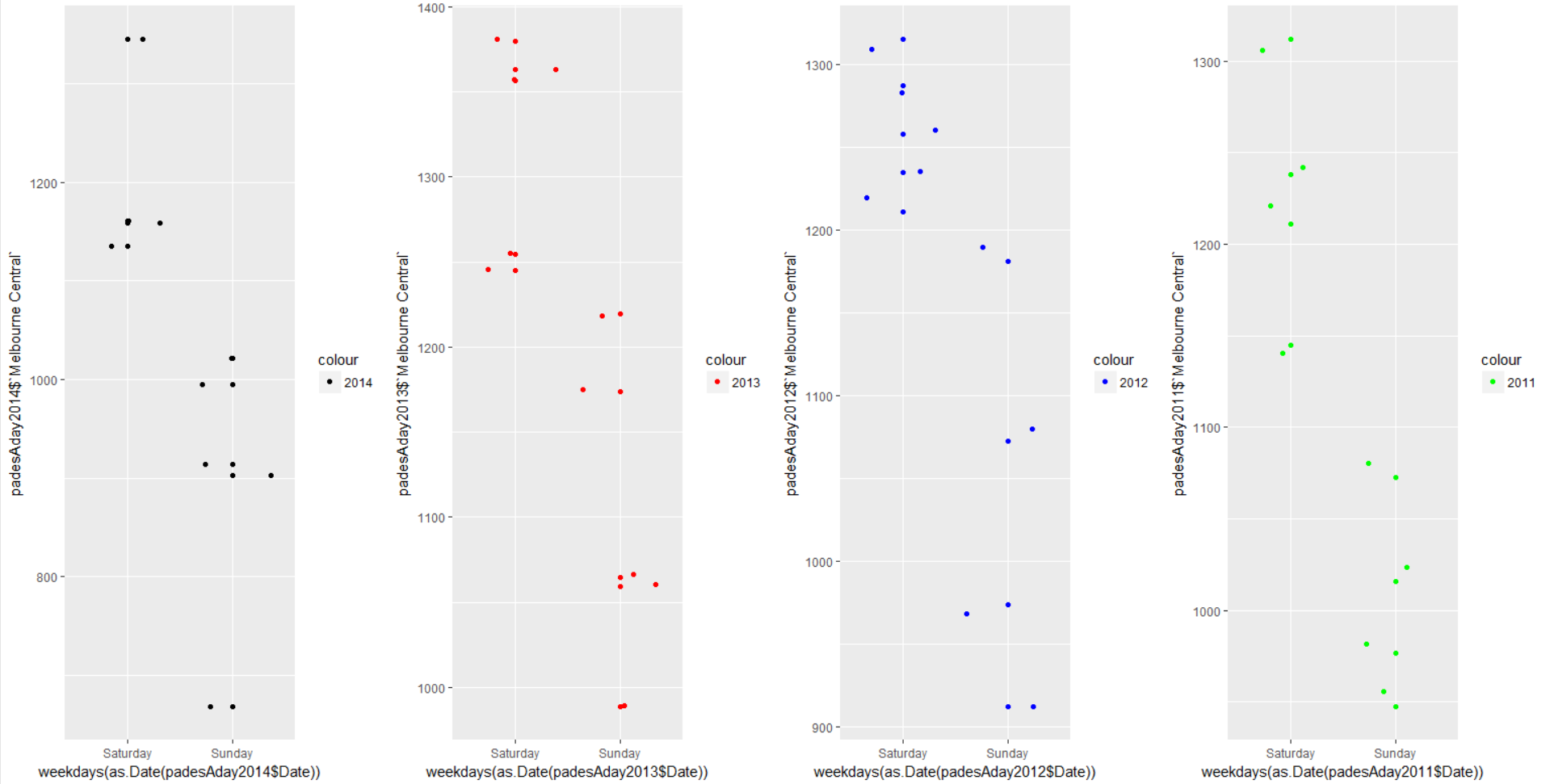


**Figure 9 Connect Tableau and MongoDB database**

After sign in figure 9, the database MelPedTra in MongoDB was connected with Tableau.

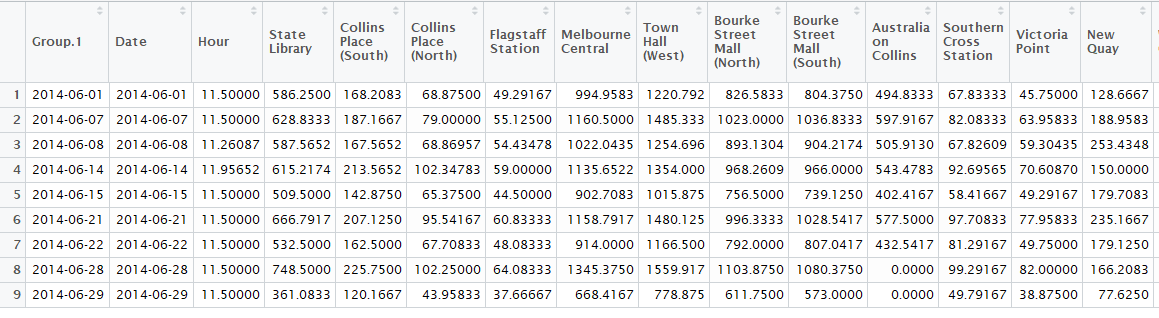
**Part 2. Statistics and Charts**

The target of this research is to find out the extension of business circles and the commercial vitality in Melbourne from 2009 to 2014. In this assignment, I only choose the data in June from 2011 to 2014 because June is the last month of a fiscal year in Australia and more discount will attract more people to shopping. I believe the data in June will provide the strong significant of business circles and commercial vitality. To highlight the feature, I only analyse the weekends in June.



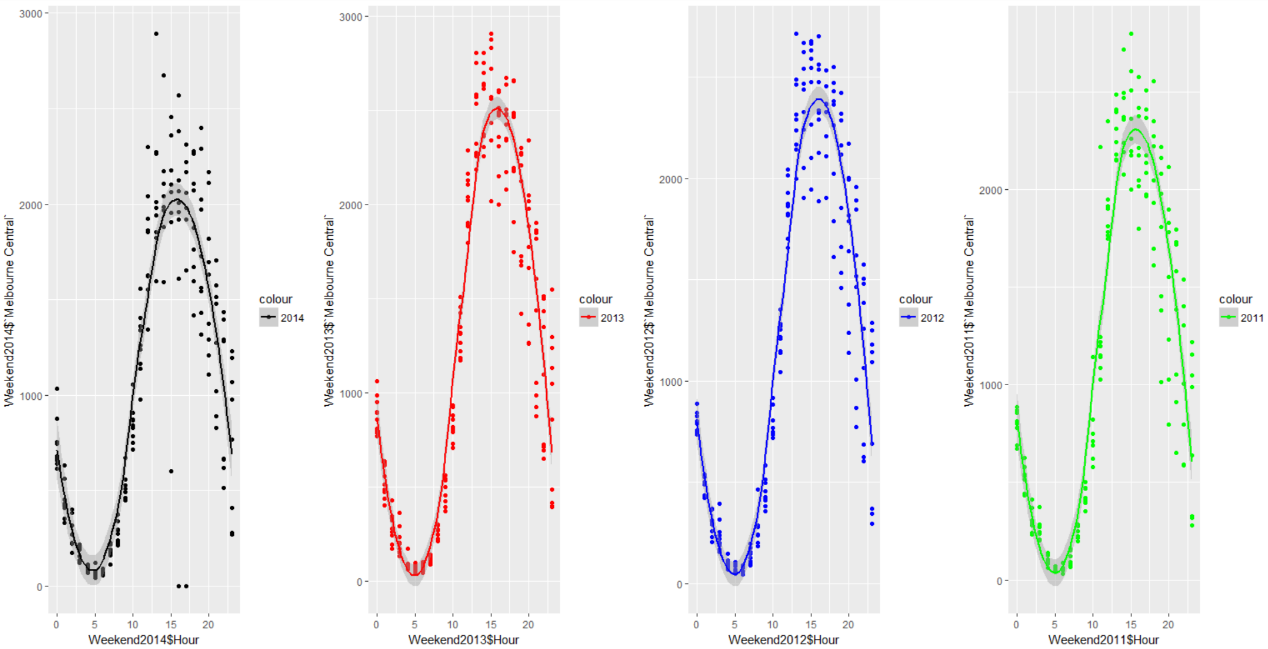
**Figure 10 Pedestrian traffic (Melbourne Central) on weekends from 2011 to 2014**

Figure 10 indicates the pedestrian traffic in Melbourne central in Sunday and Saturday from 2011 to 2014. From the figure, people prefer shopping on Saturday than Sunday. And the significant never changed in these years. Thus, in future analysis, the pedestrian in Saturday should be an important part to consider. In this figure, the mean value of a day is the marked point to compare.



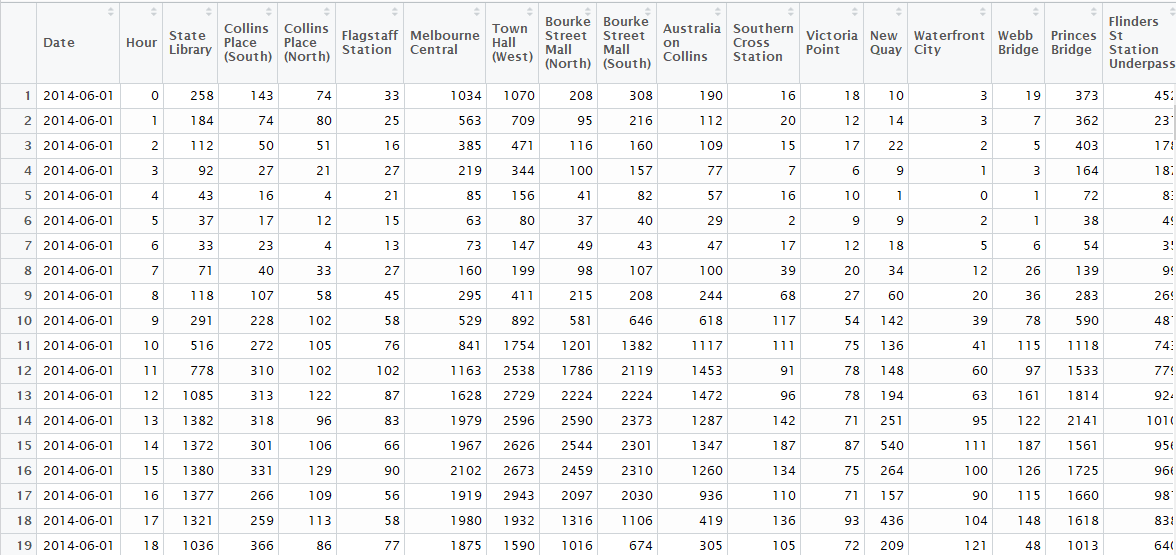
**Figure 11 Mean value in a day in 2014**

Figure 11 indicates the table of mean value in a day in 2014. This table could use to compare the density of pedestrian in deferent observation point. Combining the result in figure 10 to change the weight value of Saturday and Sunday, I believe there would be an interesting result in future research.



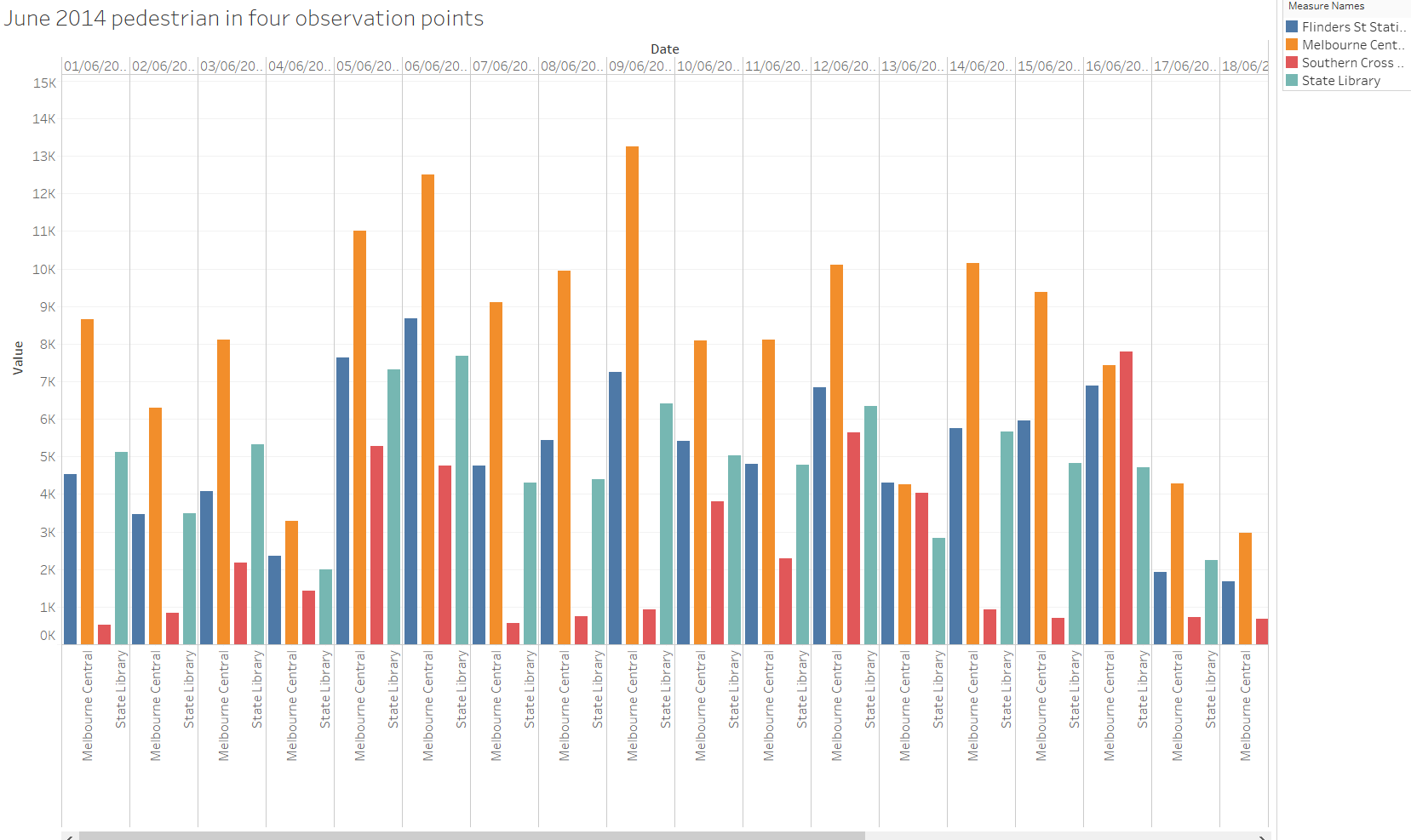
**Figure 12 Pedestrian density in a day**

Figure 12 shows the pedestrian density in a day. The fitting lines indicate the different density in different hours. The gray shadow shows the standard deviation in each hour. From the figure, the situation is almost stable because of the shadow region. Comparing the four subplots, the pedestrian density decreases in 2014.



**Figure 13 Pedestrian density in weekends**

The plotted result comes from figure 13 which indicates the pedestrian density in weekends and the table remains the density in each hour.



**Figure 14 Pedestrian density of four stations in 2014**

Figure 14 is made by Tableau. This histogram shows the sum pedestrian in a day about four different stations, Melbourne Central, Flinders St Station, Southern Cross station and State library. From the figure, people prefer to go shopping in Melbourne Central rather than reading in State library which is beside of Melbourne central. Southern Cross station always has least pedestrian which may find out that people prefer the shopping center which is near they staying.

In the future, those conclusions should be improved by more data. If the conclusions are persuasive, it could give a guide of business extension in future.