Bus Data Study

Shikun Ding, Wonyong Ha, Luit Saikia

The goal of our project largely aims to understanding IU bus data patterns of time of traveling, the number of passenger, driver working hour and related weather data. Data types and ideas were how such as rain, snow/ sleet affect bus running and passenger number, how many people using bus data in different time, and how the number of passenger change by month/day (in considering of semester begin, breaks, ending, drop class etc.

With more than 10000 riding each day, IU buses play an important role of IU transportation. Bus resource is limited and costly in real life. By understanding how bus data runs, we can optimize bus data schedule such as when very few people taking bus, decrease the number of running bus and when all buses are full or many passenger taking buses then we can increase the number running busses so that bus service can better serve the passengers.

Background

IU Bus Company responsible for providing Indiana University bus service. IU students, staff, faculty are free from charge to take IU bus. In average, more than 10000 ride using IU bus every day. In our project

The data we that have is for 2015 Spring semester and includes and some important notes: The time travel between two bus stops and stopping time at certain bus stops in seconds. At what time there are rainy dates or snow dates or sleet (rain-snow mixed) dates. We will further study what is the difference between official shift calendar and actual shift time (actual arrive bus stops' time). While the number of passengers vary by each day, month and per shift, we will study how they are different.

Different bus route has different location and carries passengers from different places. However, there are still some overlapping places. What is the difference? Number of passenger (in terms of average or by dates) for all buses for certain route, such as route A, at certain time period. Numbers of passengers also vary by drivers. Because IU bus count the

number of passenger by driver entering in number each time passenger get in to bus, they may make mistakes which means if many people come in at same time, bus driver may miss count the number. Thus, we will study what's the correlation between the number of passenger and drivers' working hour (clock in and clock out).

As a previous works related to bus data study, according to Furth (Furth, 2010), data errors can happen while collecting data. In our case, some bus drivers might over count or undercount the number of passenger because the bus driver counts the number of passenger manually. And, the time for bus running might be wrong if bus broken or during fixing then the number in seconds will be great. In one case, there is a time record show that the time is 14388081 seconds which is valueless because this number may represent bus broken or in fixing.

Second, by analyzing the bus data about running time and route level schedule adherence, we can study how buses run. However, many factors need to be considered. For example, extreme weathers such as snow or rainy might affect the number of passengers and/or when they take bus rides.

Third, more variety of types of data are needed. This means, limited data types limits analyze results. And more commercial bus data should be considered and released.

Furthermore, basic knowledge about how bus runs is needed. For instance, there are fewer passengers during weekend than school days (Monday to Friday). Less people are taking bus during the spring break. Actual arriving time might be different than scheduled arriving time. E

On the other hand, according to Lampkin and Saalmans' study (Lampkin and Saalmans), generating bus schedule based on the number of passenger vary by time can optimize efficiency.

Methods

Describe data and how we obtained it,

We obtained the data from professor Predrag Radivojac, IU bus (Logan as representative)

and IU bus website.

The data we that have is for the 2015 spring semester and includes:

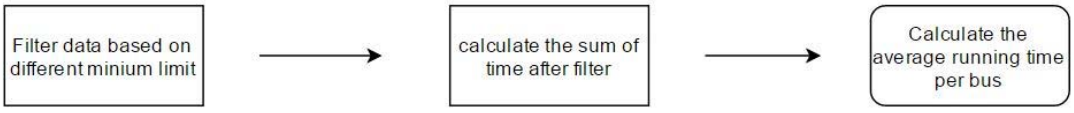
1. The time travel between two bus stops and stopping time at certain bus stops in seconds.
2. Rainy dates.
3. Snow dates.
4. Sleet (rain-snow mixed) dates.
5. Official shift calendar.
6. Actual shift time (actual arrive bus stops' time)
7. Number of passengers each day, month and per shift.
8. Number of passenger (in terms of average or by dates) for all buses for certain route, such as route A, at certain time period.
9. Number of passenger by each driver.
10. Drivers' working hour (clock in and clock out).

Describe your methodology: give flowcharts, diagrams or formulas where appropriate.

The project can be separate into several parts. For part one, time that not fit for requirement excluded out. This means, for example, I set limit to be 1000 seconds, then any time value above 1000 will be excluded. By doing this, we can exclude abnormal situations such as bus broken or fixing time. Because usually the bus travel between two bus stops will be no longer than 1000 seconds in this city. Second step is do sum of time after filter out. And finally calculate average running time per bus by the formula below:

Sum Of Running Time/bus Number = average Running Time.

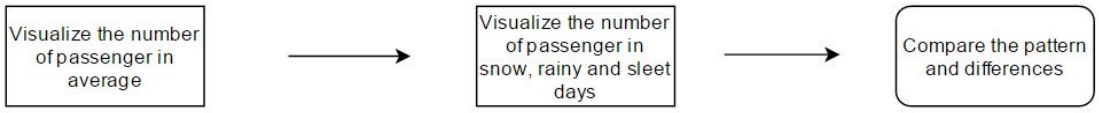
See the flowchart below for more information.



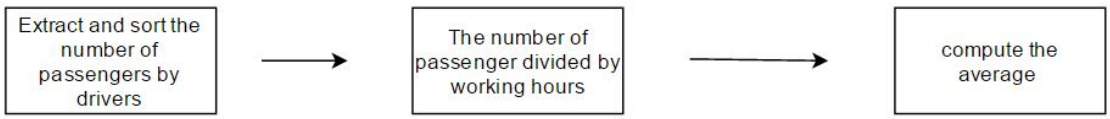
For part two, we using method that grab data about how many passengers come in to

busses group by different route at different time, in addition, visualize the result. This means, for example, we extract all buses for a certain route and match the total passenger for these busses with different time so that we know how many people get in/use the bus at certain time. We use this method to analyze four routes (A, B, E, X). Furthermore, we are able to know the max using time. This means, at what time the most people get in to the buses.

For part three, we compare the number of passengers on rainy, snow, sleet and other days. See the flowchart below.



For part four, we extract and sort the data about the number of passenger by different drivers. For example, the first driver has been taking the most number of passengers and the last driver has been taking the lease. On the other hand, we calculate the number of working hours for each driver. Then calculate in average, for each driver, how many passengers can a driver take during a typical working hour? See the formula below. See the flowchart below.
Average Taking Passenger = Sum Of Passenger Number Each Driver / Hours Of Work Each driver



Describe evaluation strategy

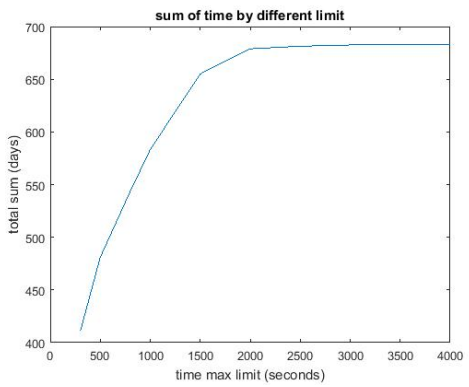
Our project largely aims at study the bus pattern. After we make conclusion and suggest to IU Bus Company, adding or decreasing bus shifts may reschedule bus schedule, observe how bus drives drive bus then we will evaluation our results bases on these ways. Data evaluation include filter out duplicate, non-valuable data and results in addition to analyze result based on actual bus running observation such as if result show spring break is most busy time for IU bus then this is impossible and mistake happens. Further evaluation for data part will be further discussed.

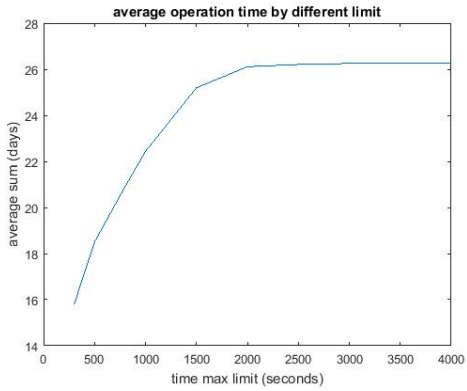
Results

According to raw data (no filter algorithm applied), the total running time (include all bus travel, wait time) is 251672197 seconds (about 4194536 minutes or 69909 hours or 2913 days).

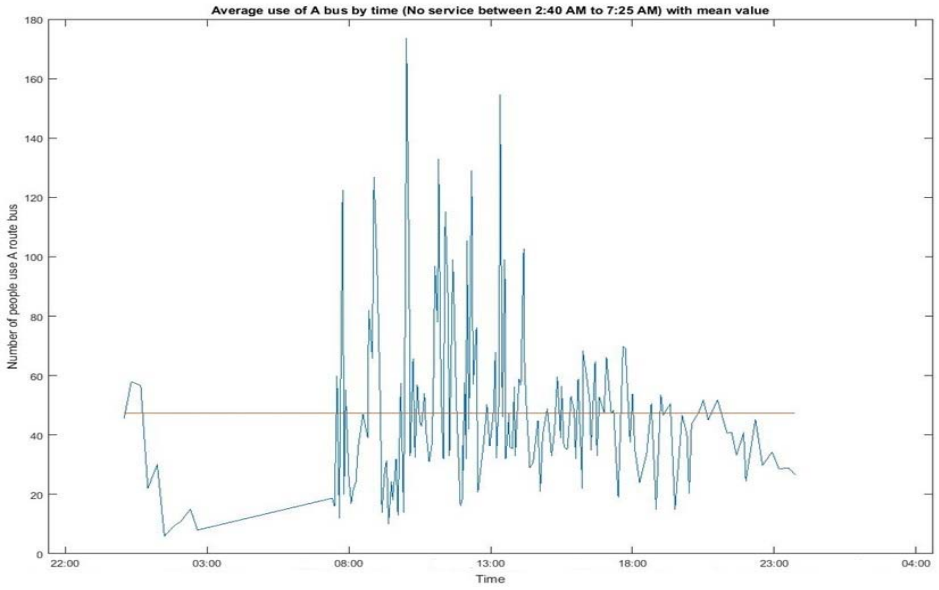
After filter, (for time < 5000 seconds), some minor bus broken down situation counts here, the total operation time is 58366742 seconds (or 675 days).

If set limit to 900 seconds which is a reasonable running time. Then each bus runs about 19.41 days in 4 months. This means, if we only count the time record that is less than or equal to 900 seconds, then each bus runs approximately 20 days for 2015 spring semester. In actual, most buses runs about 20 – 25 days (or 1728000 seconds to 2160000 seconds) for the spring semester. Left image is total running time (for 26 buses). Right image below is in average per bus.

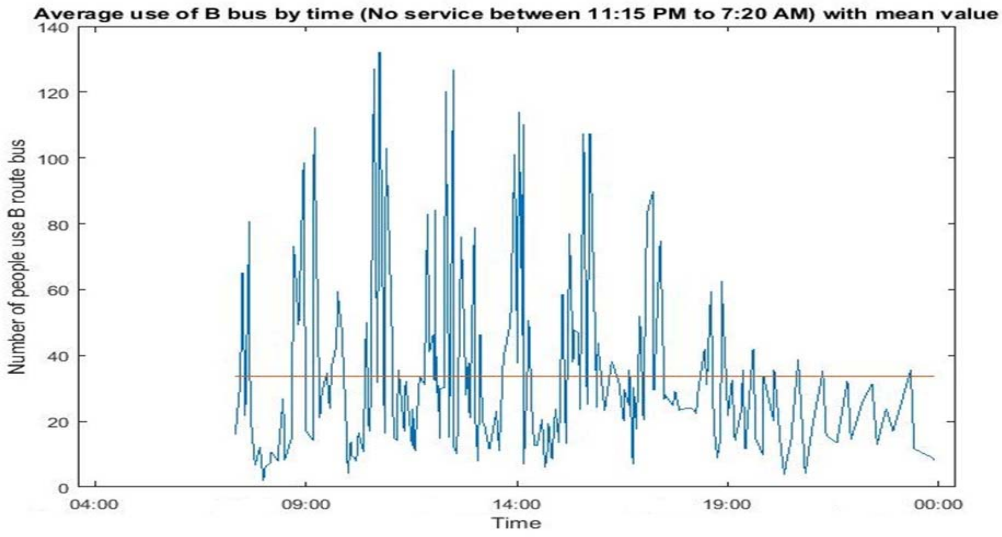




Based on the data about average use of route A, at 10:50 AM there are 173.62 people get in this route's busses. This means, for all route A buses, there are these number passenger get in. Also, the min time of passenger is at 7:50 AM, there are 6 passengers. We record the number of passenger usually each 4-5 minutes. The average is 47.46 passengers for route A busses per 4-5 minutes. The red line represents the average. Notice that there is no service between 2:40 AM to 7:25 AM. There is a line connection in the image because I'm using normal plot function. Please see the image below. Busy time is including 7:00 PM and 12:00 PM (at noon).

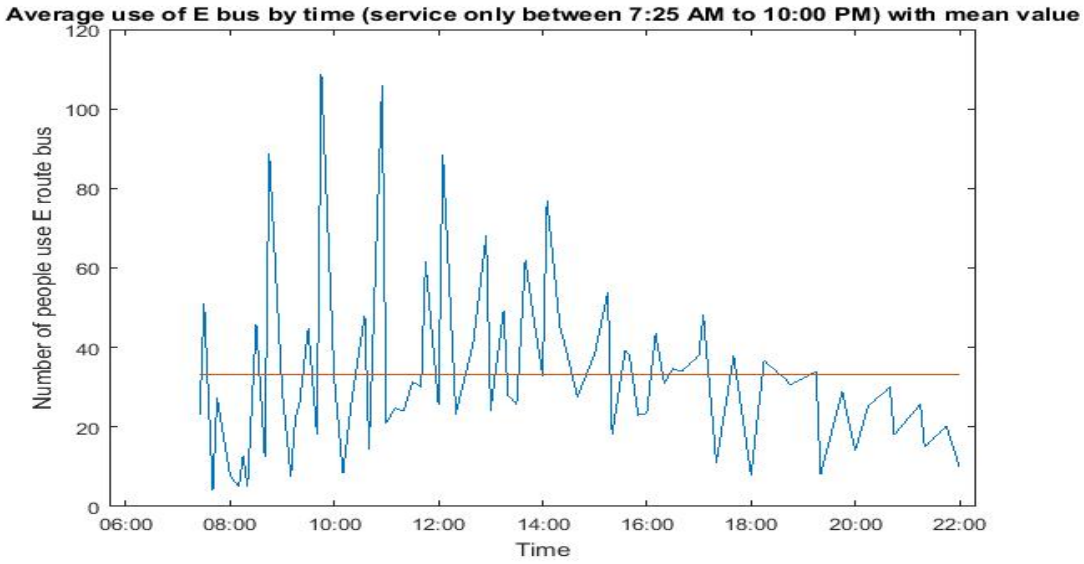


By using same method, we analyze other routes. For route B, the min passenger time is 8:00 AM, 2 passengers, which is early morning but increasing quickly. Max time is 10:45 AM with 132.27 people. The average is 33.57 passengers (in average per 5 minutes). See the image below. Busy time including 12:00 PM.

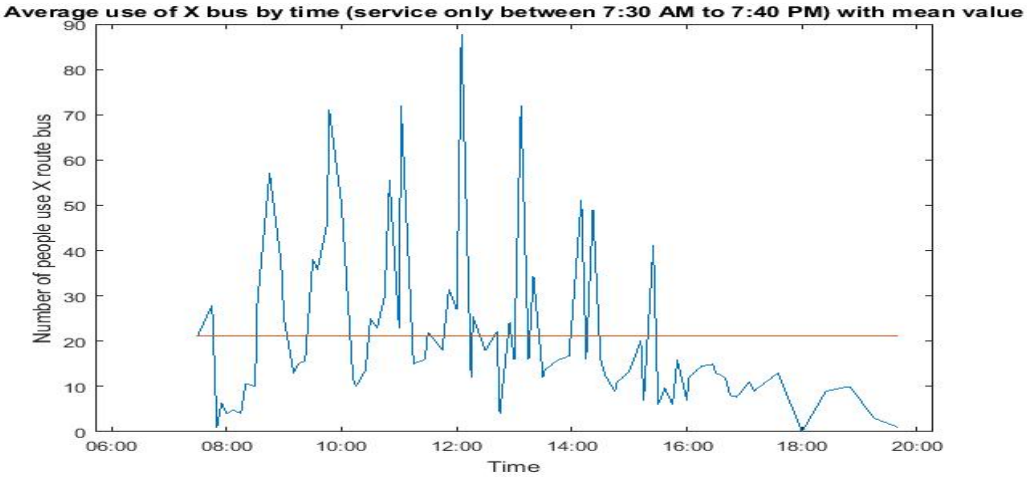


For route E, the max is 9:45 AM with 108.8 people. Min time is7:40 AM for 4 passengers. And average is 33.19. One busy time is 10:00 AM but it is gone very quickly. This means many people take route E bus at that time but no more significant people coming

after that time. Check the image you can see that the line increase and drop very fast. Then, during 12:00 PM to 17:00 PM and 18:00 to 19:00, there is constant number of passengers means people constantly coming but not crowded. See the image below.

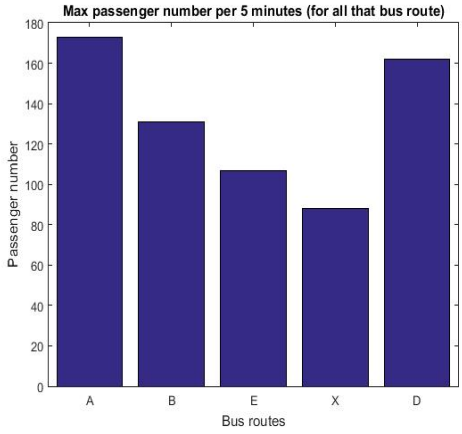


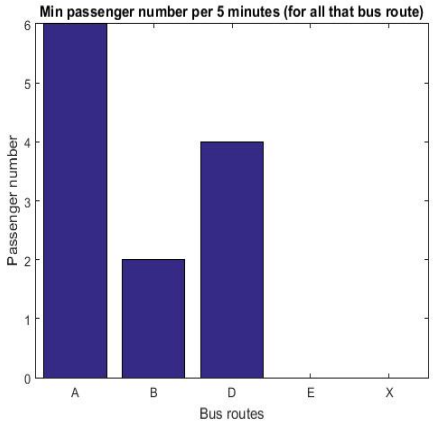
For route X, the max time is 12:05 PM with 87 passengers. Min time is 18:00 PM with 0 people and the mean is 21. Busy time is 9:00 AM. Again, all red lines here and above means average.

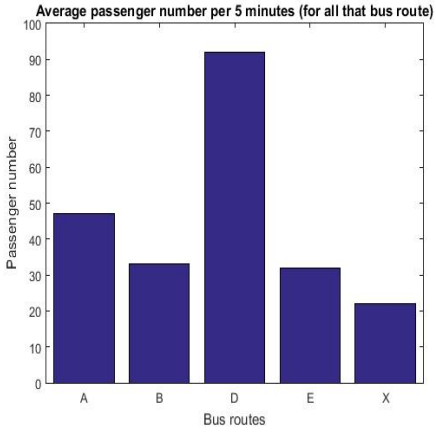


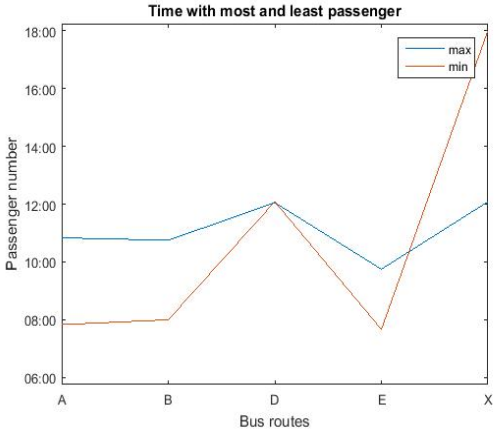
Here are some more visible bar plots about max and min passenger data. Notice that we only have limited (less than 10 data points) for route D. Thus, result for route D is less

considerable.





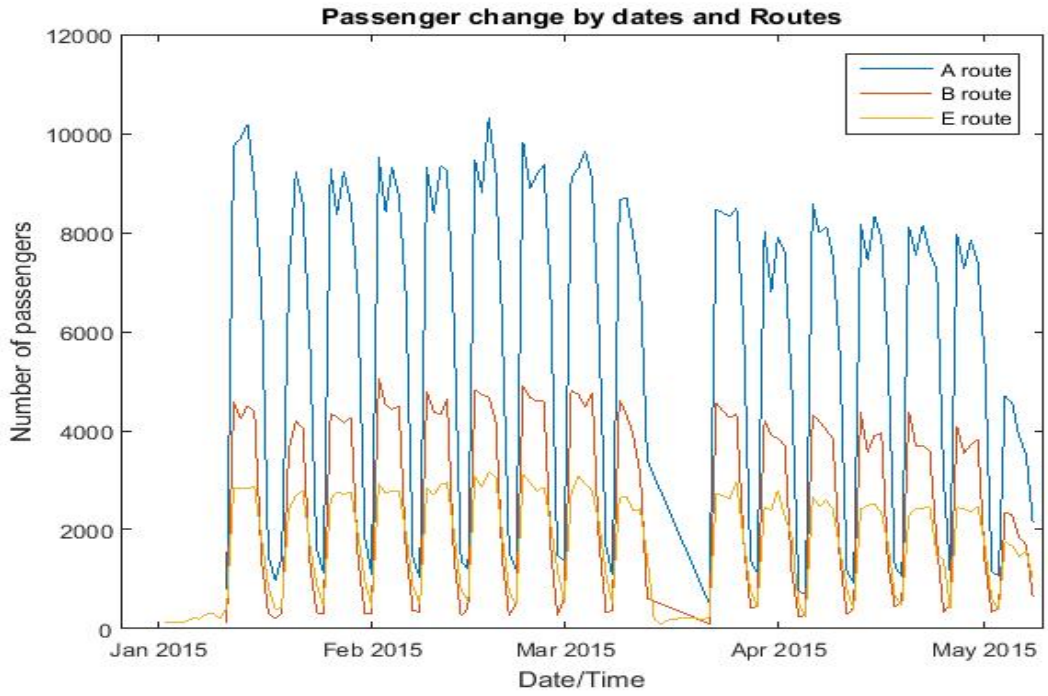




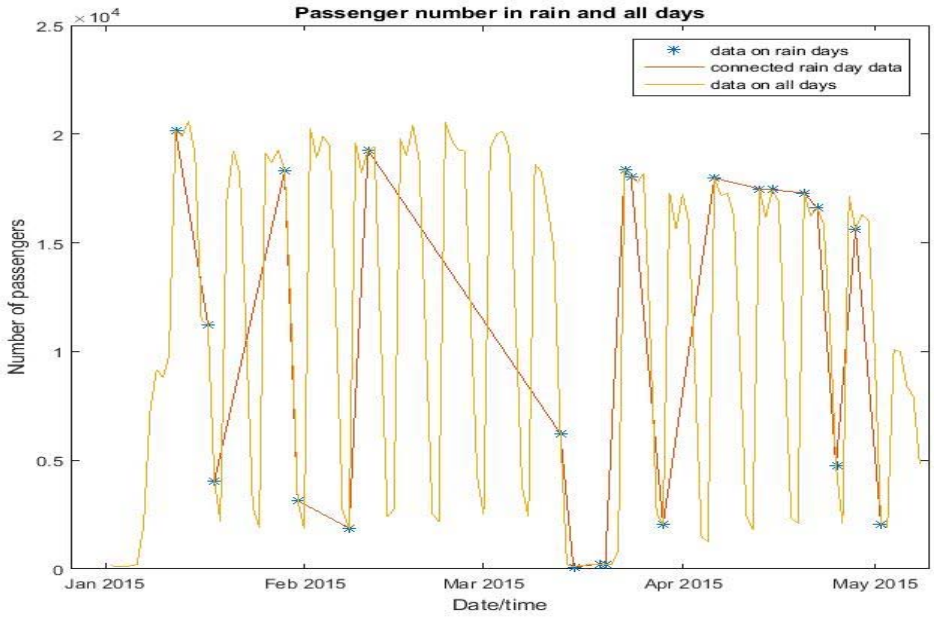
Next part, we analyzed the pattern about how passenger number change by dates. Bus routes show very similar pattern. The graph below show the passenger change by dates and routes. The peaks represent school days (Monday-Friday) and bottom represent weekend or breaks. Notice that more passenger A route, than B route than E route. So, route A is the most popular one and E is the least popular route.

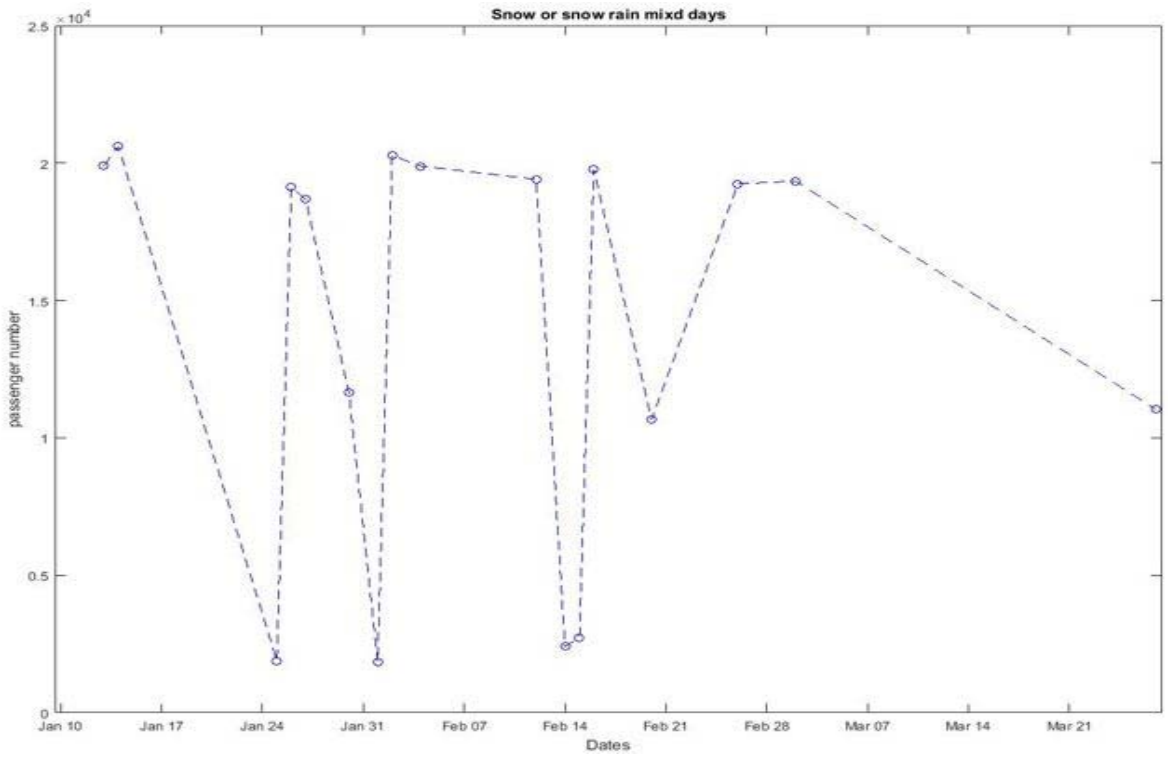
Furthermore, March 14 to March 22 is during spring break, and very few people taking bus during that time. Our graph proved this idea. On the other hand, the numbers of passenger start to decrease after week 8 and week 9 because students need to drop class during week 8 (for 16 weeks class) and week 9 (for second 8 week class).

And, the number of passenger decreased about 40 % after beginning of May because many people finished final exam and start leaving. See our graph below.



After we analyze the number of passenger in different weather, rainy weather usually not affect the number of passenger but snow or sleet (snow and rainy mixed) weather cause the number of passenger increase about 30%. In snow or sleet days, there are 3044 more than all days, in average (27.71%) and 3463 more than only rain days, in average (32.77%). I will discuss this more in the conclusion part. See the graph below for more details.

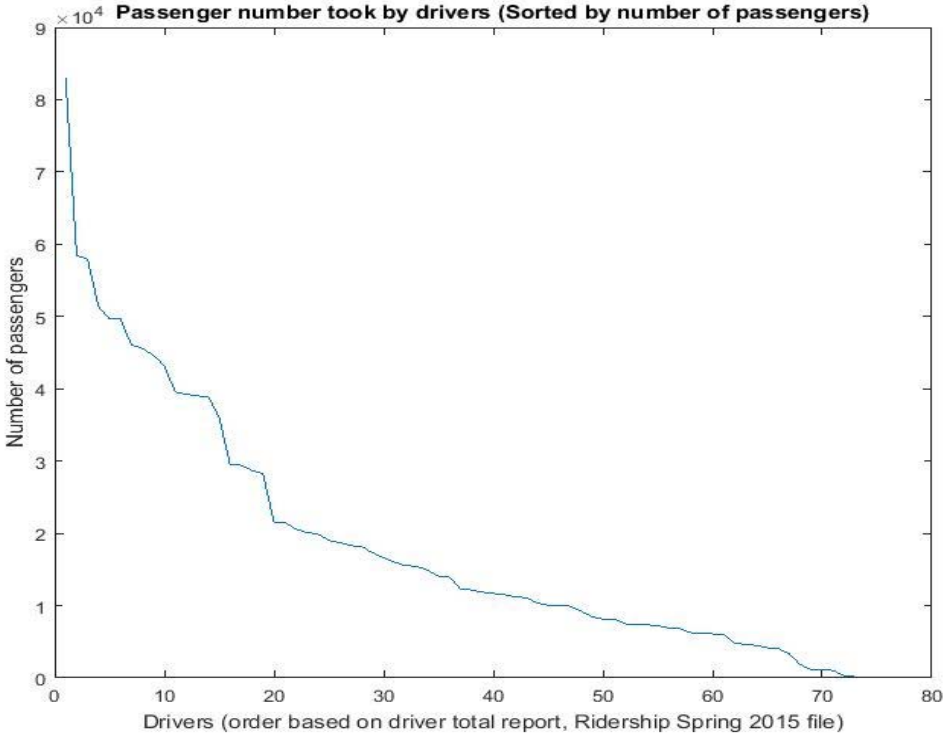


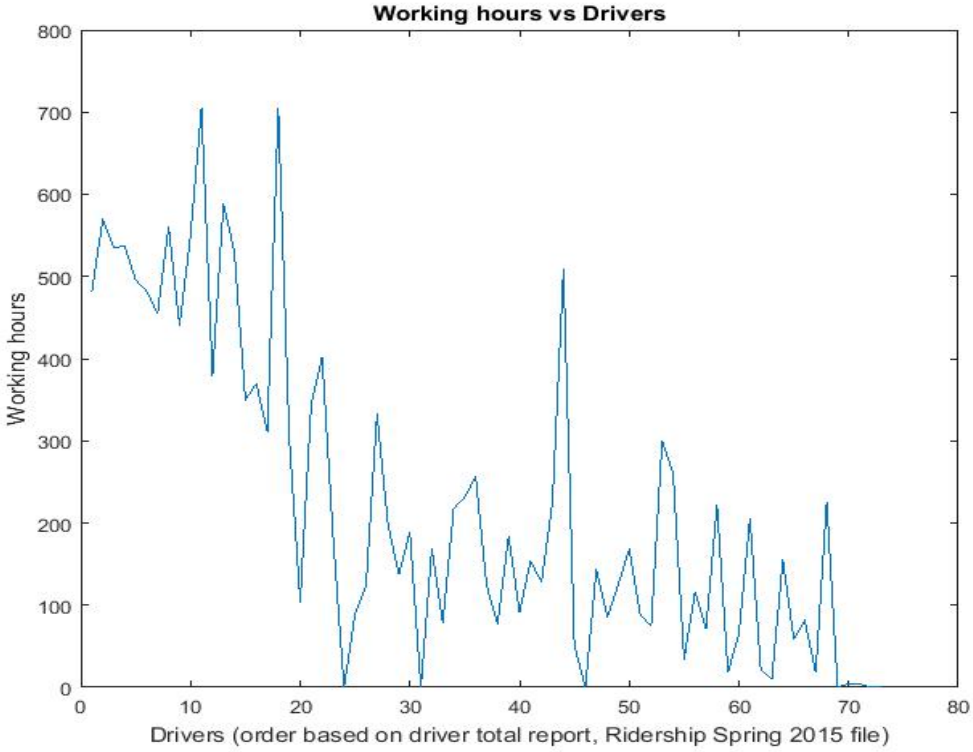


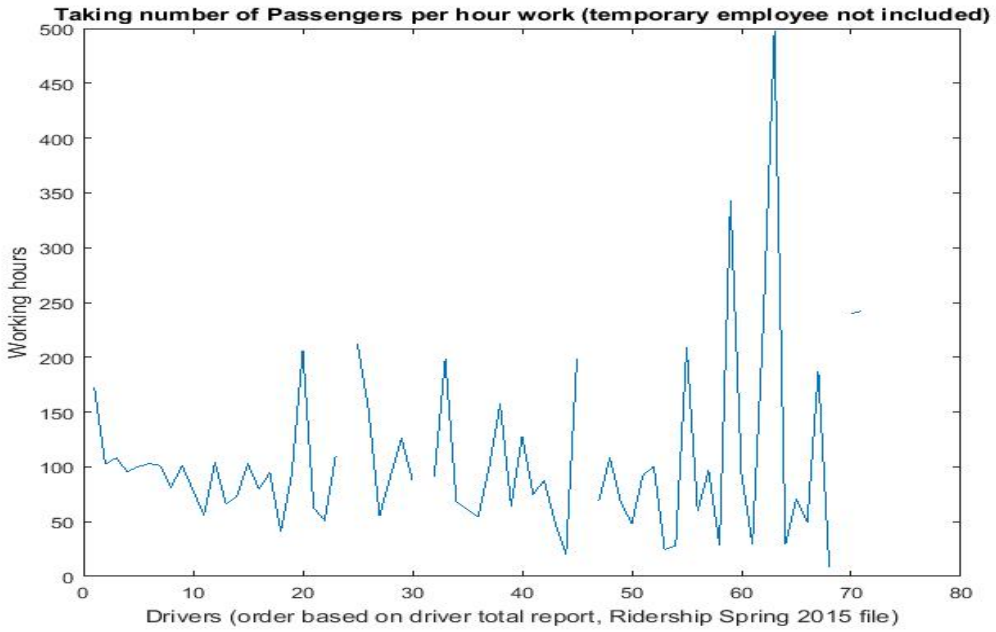
For the part about driver data analysis:

1. Driver K. 497.6344 passenger per hour in average. Clocked in 9:18:00 and took 4628 passenger (May be over count passenger)
2. Driver M. 8.0986 passenger per hour in average. clocked in 225:06:00 and took 1823 passengers (May be under count passenger)
3. For all drivers, there are 1395086 passengers. Each driver took 19111 passenger
4. In average, drives worked for 7220 hours. 1395086/7220 = 193.2252. Thus, in average, drivers can take 193.22 passenger during 1 hour of work.

For the first image below, x-axis represent the drivers. The first driver counted most passengers and the last (as shown on the most right of image) driver count the least passenger. For the third image, the gaps (unconnected line) represent temporary employee that no working hour data found.


Passenger number took by drivers (Sorted by number of passengers)
×10^4
Number of passengers
Drivers (order based on driver total report, Ridership Spring 2015 file)
0
1
2
3
4
5
6
7
8
9
0
10
20
30
40
50
60
70
80


Working hours vs Drivers
Working hours
Drivers (order based on driver total report, Ridership Spring 2015 file)
0
100
200
300
400
500
600
700
800
0
10
20
30
40
50
60
70
80



Conclusions

Each bus runs about 19.41 days (465.85 hours) for spring 2015 semester, exclude abnormal time such as broken, fixing and extreme traffic issue. Actual data show 19 to 25 days of running time in average of each bus.

Average number of passengers per 4-5 minutes: A: 47.46, B: 33.5, E: 33.1, X: 21.1461.

IU bus company need to consider add bus shifts during busy time. The busy time for different bus routes are:

A route: 7:00 PM and 12:00 PM (at noon).

B route: 12:00 PM

E route: During 12:00 PM to 17:00 PM and 18:00 PM to 19:00 PM

X route: 9:00 AM

The numbers of passenger start to decrease after week 8 and week 9 because students need to drop class during week 8 (for 16 weeks class) and week 9 (for second 8 week class). And, the number of passenger decreased about 40 % after beginning of May because many people finished final exam and start leaving. See our graph below. So, **IU Bus Company need to consider decrease bus shift after week 9, we suggest about 10%. And, decrease**

the bus shift during final's week (week 16); we suggest 30 %.

On the other hand, rainy weather usually not affects the number of passenger but snow or sleet (snow and rainy mixed) weather cause the number of passenger increase about 30%. In snow or sleet days, there are 3044 more than all days, in average (27.71%) and 3463 more than only rain days, in average (32.77%). There is several reasons cause more passengers in snow weather. First, Many students who has car, tend to take bus in snow weather cause their car can't run in extreme cold weather. Second, Low base cars (small cars) can't run in heavy snow weather. (SUVs and buses with higher base level can run). And, it is harder for people to walk in snow than rain/sunny days. Thus, **we suggest add more buses to serve during snow (or sleet) weather.**

Future work for this bus data include finding out who is the best driver by analyzing arriving bus stops on time. How temporary driver's drive patters (counting passenger, arrive on time) different than other drivers. And calculate bus run efficiency in different time. (Number of passenger/how many bus run at that time).

References

Furth, P. (2000). *Data analysis for bus planning and monitoring*. Washington, D.C.: National Academy Press.

Lampkin, W., & Saalmans, P. (n.d.). The Design of Routes, Service Frequencies, and Schedules for a Municipal Bus Undertaking: A Case Study. *Or,* 375-375.

Indiana University Campus Bus. (n.d.). Retrieved December 8, 2015, from http://www.iubus.indiana.edu/campus_bus/bus_schedule.html