

I am going to analyse average number of live births in Singapore from 2005 to 2016 after aggregating the function and plotting a scatter plot (Average live births vs Year) as well as a boxplot.

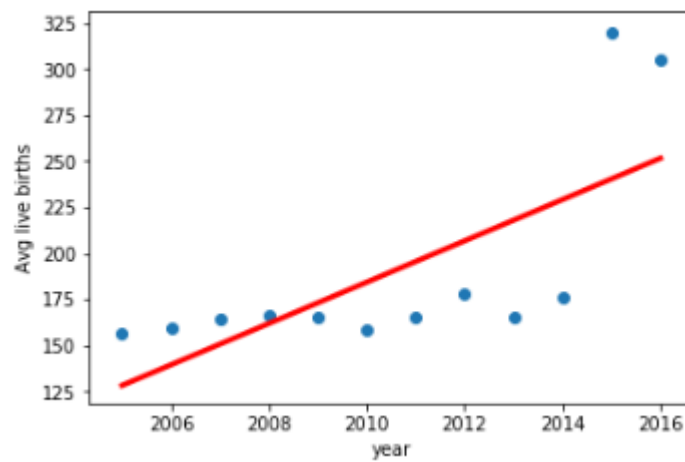


Figure 12: Average live births (scatter plot, linear regression)

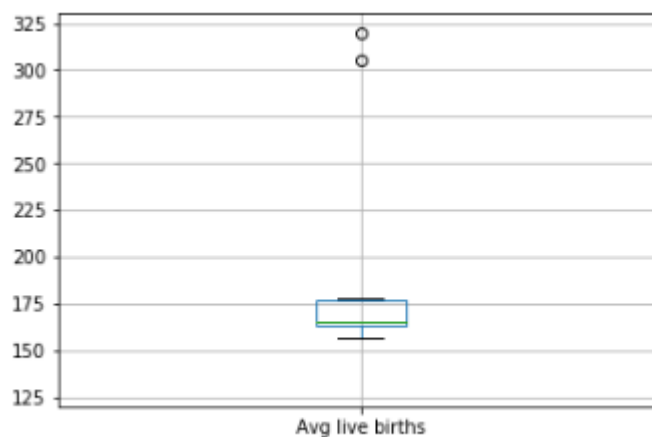


Figure 13: Average live births (boxplot)

By looking at Figure 13, there are two outliers as denoted by the two circles. These two outliers correspond to the year 2015 and 2016 in Figure 12 (scatter plot).

In figure 12, the linear regression is increasing for every year. This means that average live births increase for every year.

I want to predict the average live births for the year 2020 based on the linear regression model. The slope of the model as shown in Figure 12 is 11.208 (rounded up to 3d.p.) and the intercept of the model is -22344.764 (rounded up to 3d.p.). Using the formula $Y = m \cdot X + C$ where Y is the average live births and X is the year, we

would be able to get the average live births for 2020. Substitute the following numbers in the function appropriately.

$$Y = m \cdot X + C$$

$$Y = (11.208 \cdot 2020) + (-22344.764)$$

$$Y = 295.396$$

Therefore, the average live births for the year 2020 will be 295 passengers.

However, the linear fit as shown in Figure 12 does not look to be a good fit to me because almost all of the data points do not appear on the regression line. Therefore, the prediction for 2020 may not be accurate.

Hence, I replotted the regression model as shown in Figure 14 for the scatter plot and came to a conclusion that cubic regression model will be a good fit as most of the data points appear on the regression line. Therefore, if I were to use this regression model to predict the average live births for the year 2020, it would be accurate.

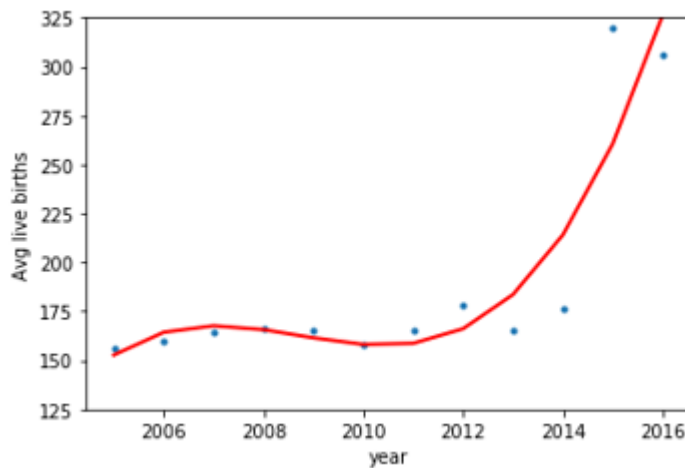


Figure 14: Average live births (scatter plot, cubic regression)

I would also like to analyse which ethnic group has the highest median of live births.

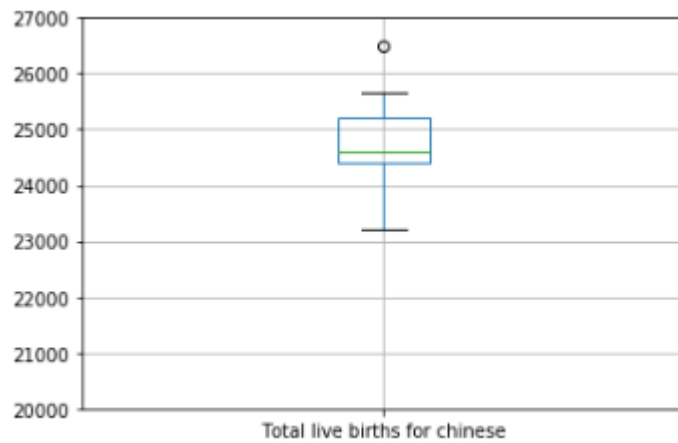


Figure 15: Total live births for Chinese

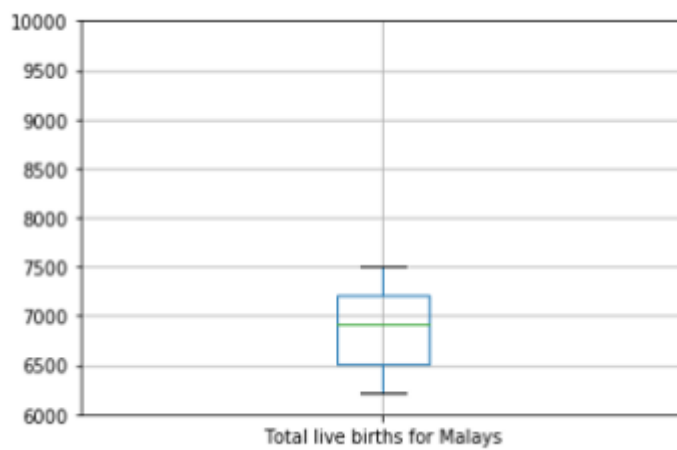


Figure 16: Total live births for Malays

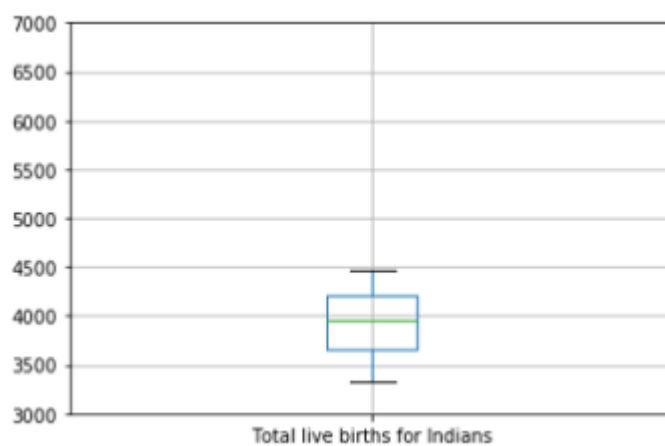


Figure 17: Total live births for Indians

By looking from Figure 15 to 17, we can tell that total live births for the Chinese (approximately 24500 live births) is the highest followed by Malay (approximately 6800 live births) and total live births for Indians (approximately 3900 live births) is the lowest. This is because the majority of the people in Singapore are Chinese (74.3% as of 2015) as shown by the statistics in Figure 18; therefore, they certainly do contribute to higher live births as compared to other ethnic groups.

Ethnic composition of resident population									
Ethnic group	1970 ^[46]	1980 ^[46]	1990 ^[12]	2000 ^[12]	2010 ^[12]	2011 ^[12]	2012 ^[12]	2013 ^[12]	2015 ^[47]
Chinese	77.0%	78.3%	77.8%	76.8%	74.1%	74.1%	74.2%	74.2%	74.3%
Malays	14.8%	14.4%	14.0%	13.9%	13.4%	13.4%	13.3%	13.3%	13.3%
Indians	7.0%	6.3%	7.1%	7.9%	9.2%	9.2%	9.2%	9.1%	9.1%
Others	1.2%	1.0%	1.1%	1.4%	3.3%	3.3%	3.3%	3.3%	3.2%

Figure 18

I would also like to analyse the highest median of live births based on the educational qualification of mothers.

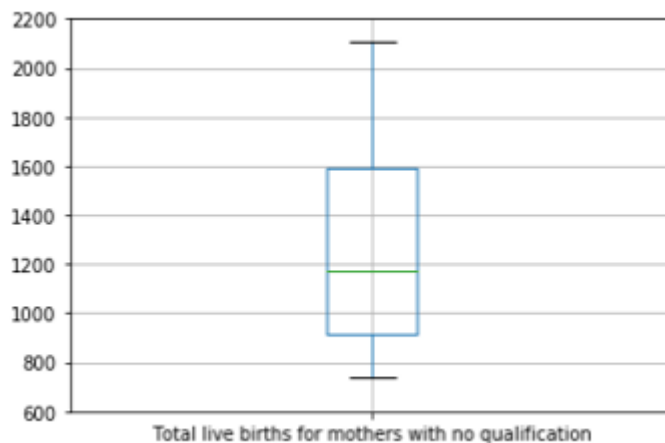


Figure 19: Total live births for mothers with no qualification

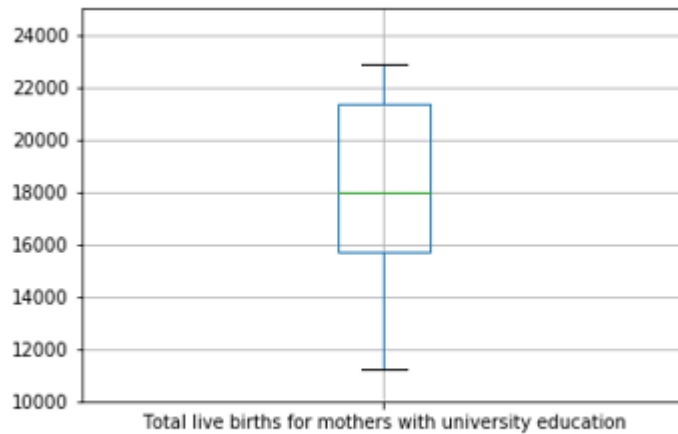


Figure 20: Total live births for mothers with university education

By looking at Figure 19 and Figure 20, total live births for mothers with University education (approximately 18,000 live births) is higher than total live births for mothers with no qualification (approximately 11,900 live births).

References:

- 1) https://data.gov.sg/dataset/live-births-by-age-group-ethnic-group-and-education-qualification-of-mother?view_id=b59686b8-b001-4087-9e5e-91868446f2d6&resource_id=257155dd-cedf-40b3-bfb4-f470f0b4cfa4
- 2) https://en.wikipedia.org/wiki/Demographics_of_Singapore#Ethnic_groups