

## INTRODUCTION TO DATA SCIENCE

### ASSIGNMENT 1

#### 1. Why are the values lower in 2016 than they were in 2015?

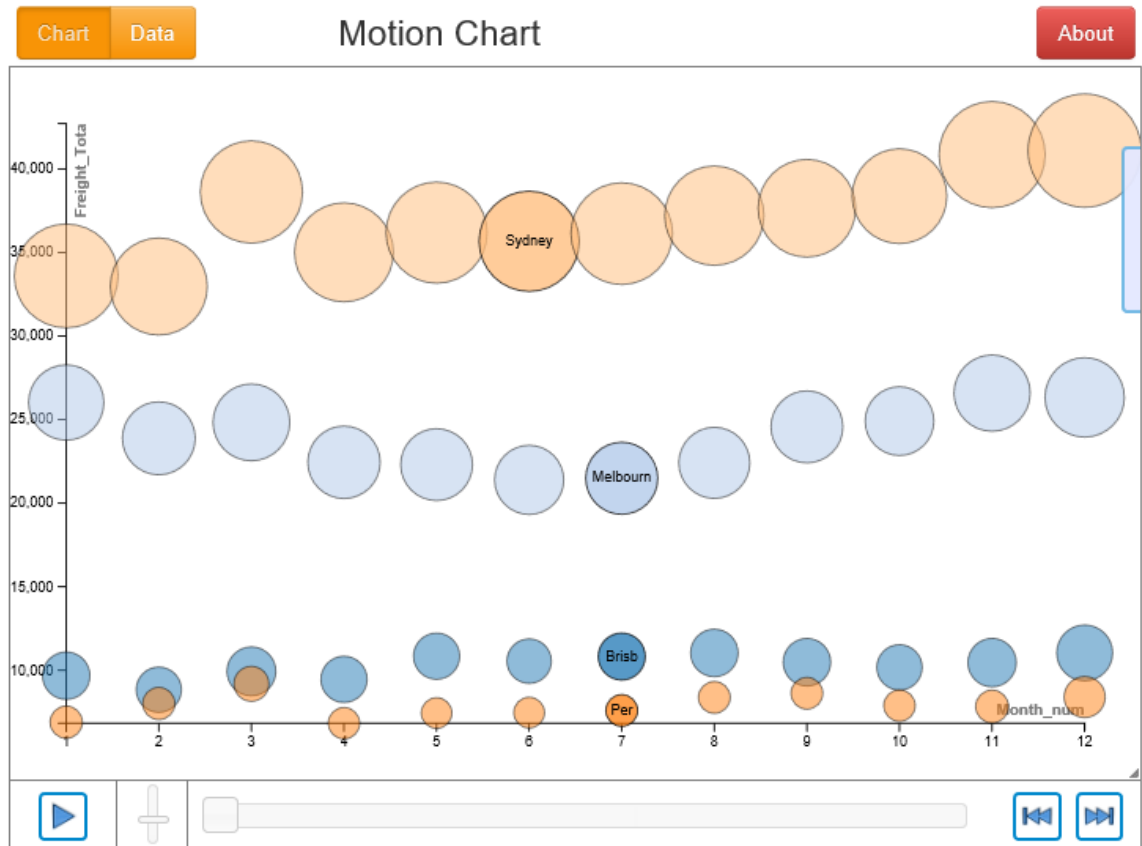


Figure 1: Visualising Airport Traffic for 2015

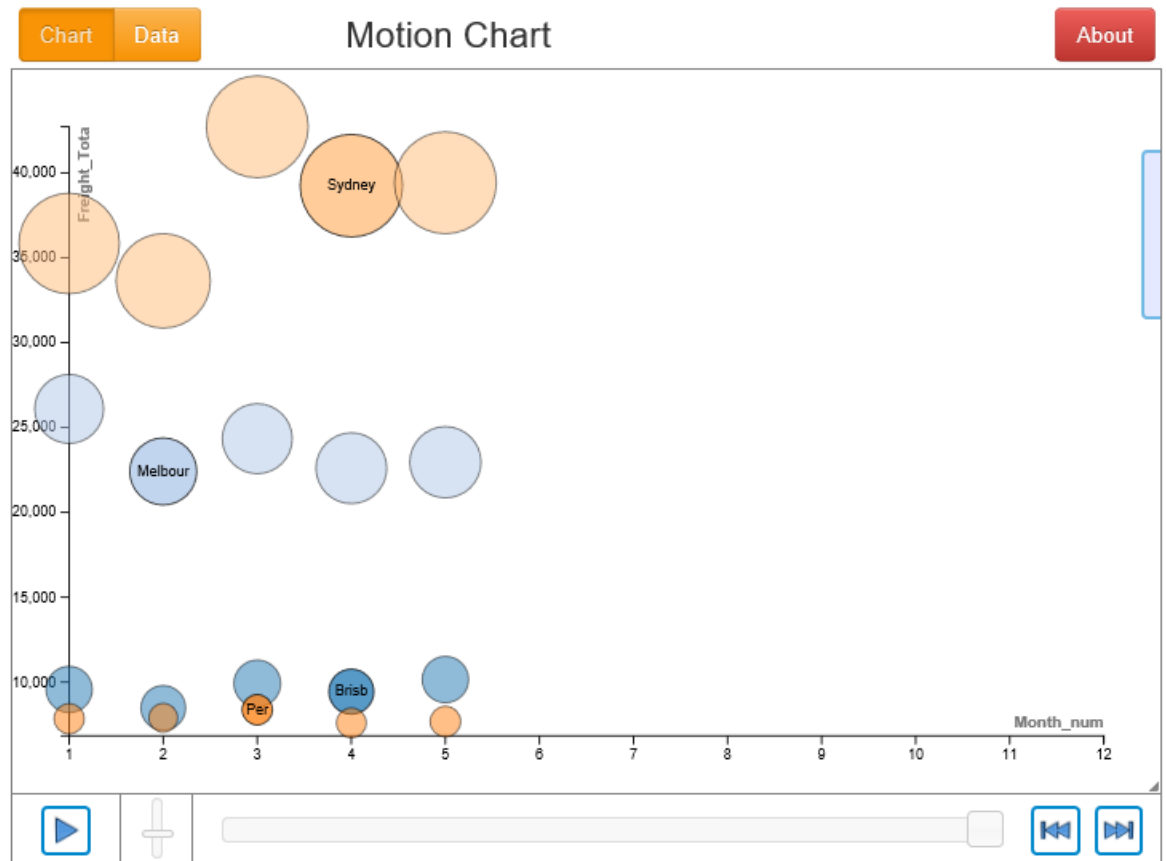


Figure 2: Visualising Airport Traffic for 2016

For 2015, it calculates total sum of passengers, freight and mail for 12 months (Jan to Dec) as shown in Figure 1. Whereas for 2016, it calculates total sum of passengers, freight and mail for only 5 months (Jan to May) as shown in Figure 2. This proves that there is incomplete data in the year 2016 since they only have data for 5 months instead of having for all 12 months. Given that we are only calculating the sum for the first five months in 2016, the total number of passengers, freight and mail are bound to be lower than 2015. This explains the reason behind the values being lower in 2016 than they were in 2015.

2. Which city has the largest number of international air passengers travelling through it in 2015?

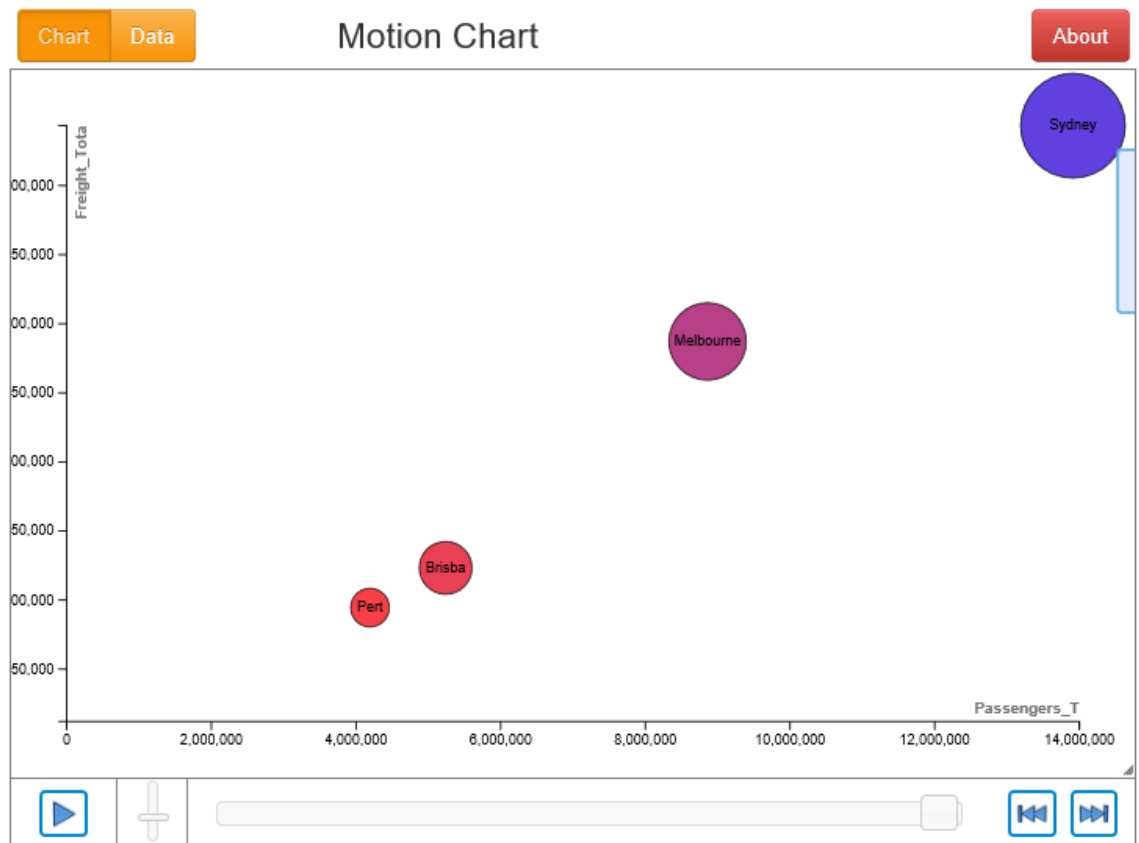


Figure 3: Visualising airport traffic in 2015

As shown in Figure 3, Sydney has the largest number of international air passengers of 13911228 passengers travelling through it in 2015.

3. In which year did Brisbane have almost the same number of passenger numbers as Melbourne?

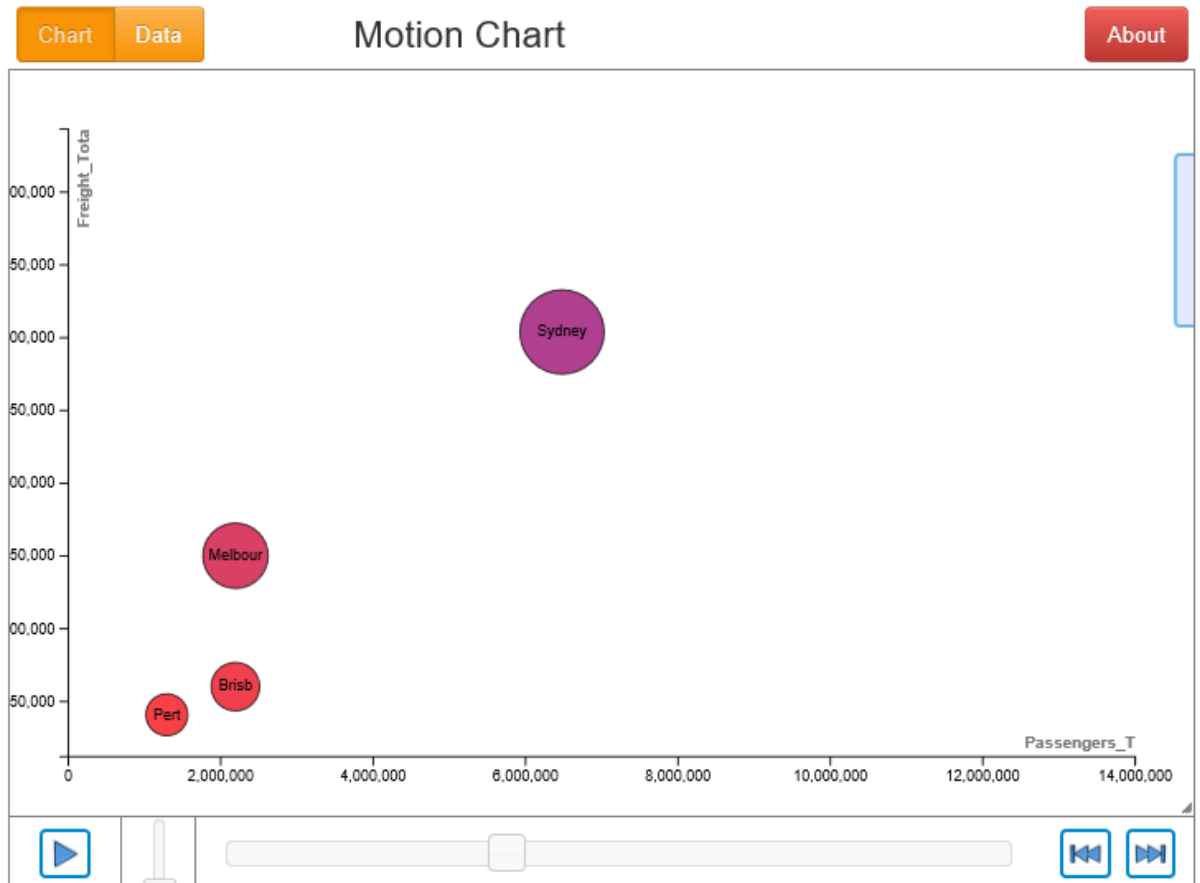


Figure 4: Visualising Airport traffic in 1996

Brisbane and Melbourne have almost the same number of passengers in the year 1996. Brisbane had a total of 2192110 passengers while Melbourne had a total of 2193309 passengers.

4. **(Bonus question)** Has the number of passengers travelling through Sydney airport ever decreased from one year to the next? If so, when did it happen? Any idea why that might have occurred?

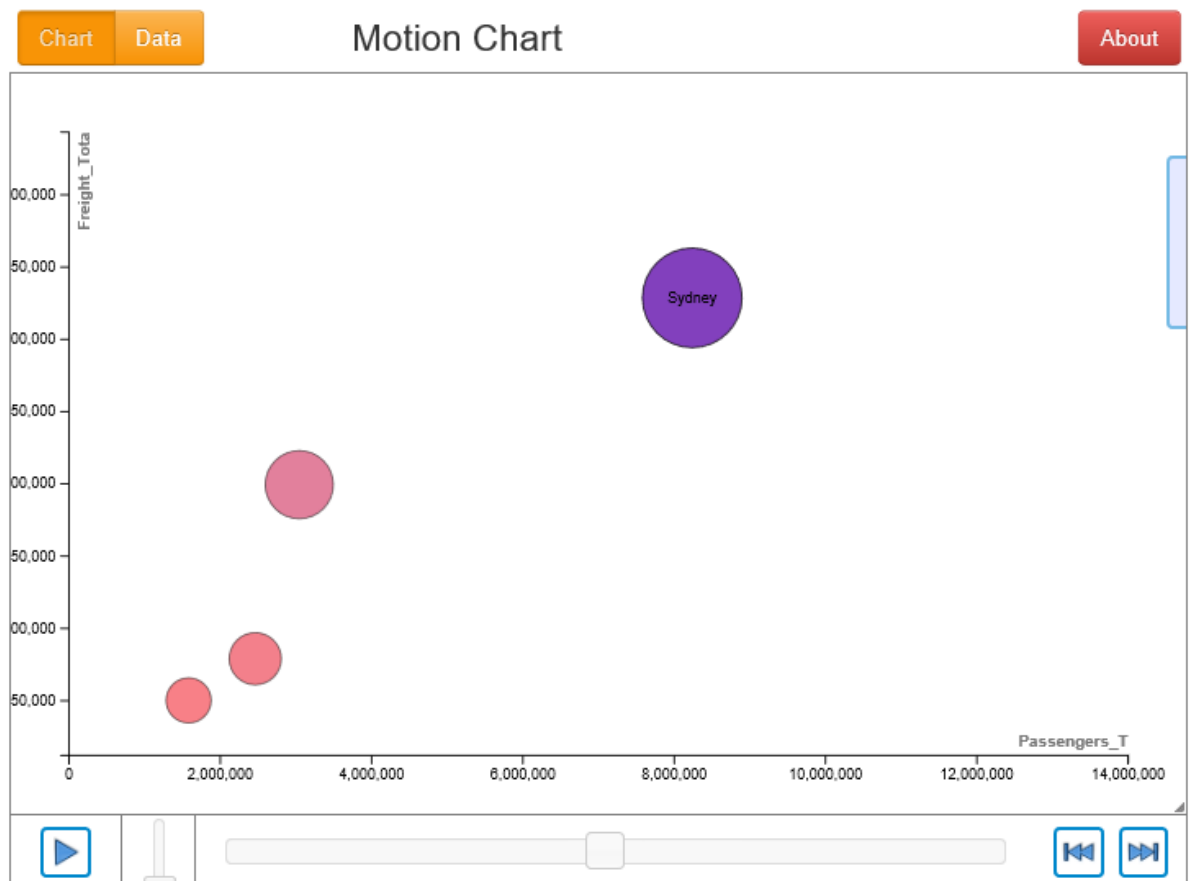


Figure 5: Visualising Airport Traffic in 2000

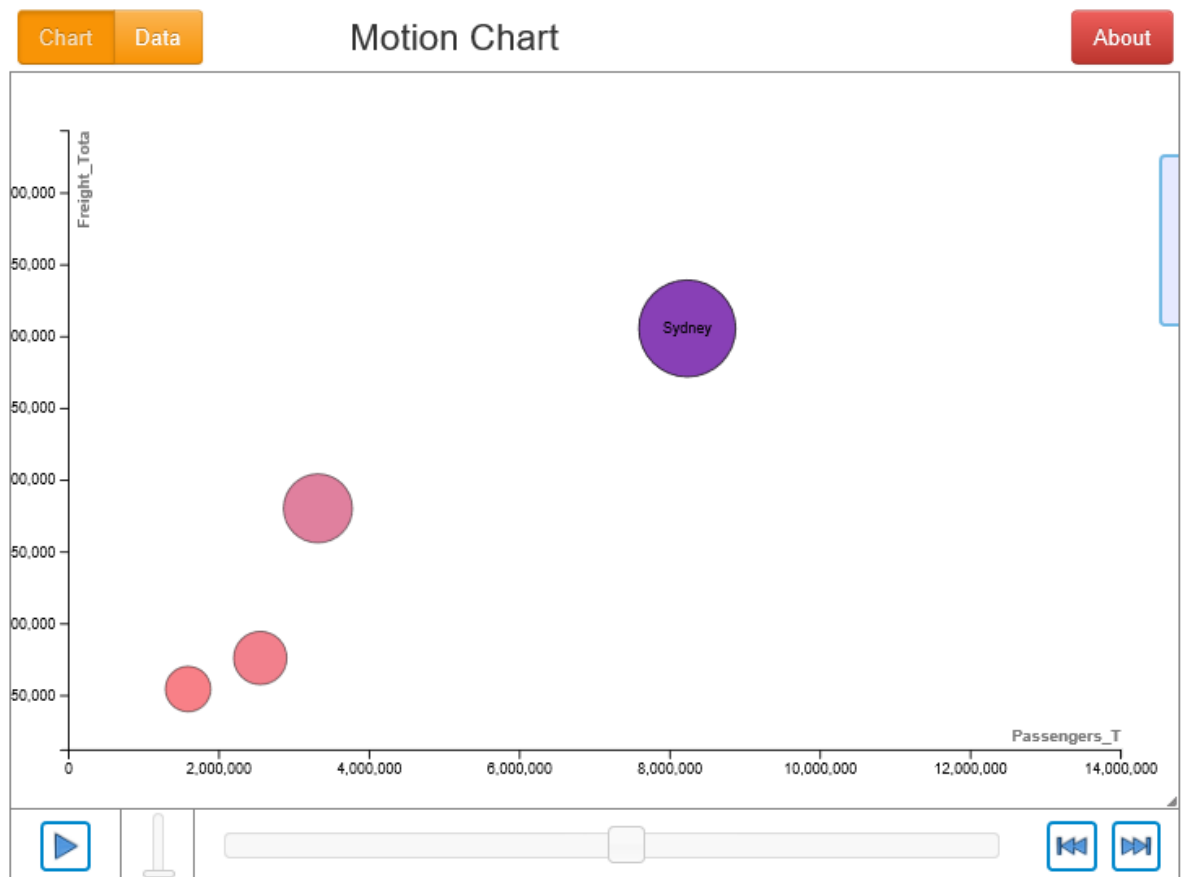


Figure 6: Visualising Airport Traffic in 2001

The passengers travelling through Sydney airport has decreased from 2000 (8237223 passengers) as shown in Figure 5 to 2001 (8228973 passengers) as shown in Figure 6. This is because in 2000, summer Olympics took place in Sydney given that Sydney was a host city. Therefore, there will be more people travelling to Sydney to witness the summer Olympics hence there would have been more passengers in the Sydney Airport in 2000. In 2001, there was no such event, hence a drop in the number of passengers visiting Sydney in comparison to the number of passengers visiting Sydney Airport in 2000.

Another reason that could have contributed to the drop in the number of passengers in 2001 could be due to the September 11 attacks that happened in 2001. Because of this attack people in Sydney might have been afraid to fly via planes to other countries. Which resulted in a fall in the total number of passengers in 2001 from 2000.

References for question 4:

<http://www.australia.gov.au/about-australia/australian-story/sydney-olympic-games-2000>

<http://www.history.com/topics/9-11-attacks>

Copy of the code to create the motion chart:

```
from matplotlib import pyplot as plt
# plt is the name we are giving to the imported library.
# We use it whenever we want to call a function provided by the library
```

---

```
import pandas as pd
# importing the library and name it as pd
```

---

```
city_reports = pd.read_csv('CityPairs.csv')
#Read csv into dataframe
```

---

This motionchart is generated to see why the values in 2016 are lower than 2015

---

```
filt2=((city_reports.Year==2015) | (city_reports.Year==2016)) &
((city_reports.AustralianPort=='Melbourne') | (city_reports.AustralianPort=='Perth') |
(city_reports.AustralianPort=='Sydney') | (city_reports.AustralianPort=='Brisbane'))
ans=city_reports[filt2]
#select rows where the AustralianPort is either Melbourne, Perth, Sydney or
Brisbane where the yr is either 2015 or 2016
```

---

```
city_class1=ans.groupby(['AustralianPort','Year','Month_num'])['Passengers_Total','Fr
eight_Total_(tonnes)','Mail_Total_(tonnes)']
sample_data1=city_class1.sum()
```

---

```
from motionchart.motionchart import MotionChart
import pandas as pd
```

---

```
%%html
<style>
.output_wrapper, .output {
    height:auto !important;
    max-height:1000px; /* your desired max-height here */
}
.output_scroll {
    box-shadow:none !important;
    webkit-box-shadow:none !important;
```

```
}  
</style>
```

#the following html block code is to ensure that the entire motion chart can be seen nicely in the output cell

---

```
mChart = MotionChart(df = sample_data1, key='Year', x='Month_num',  
y='Freight_Total_(tonnes)', xscale='linear', yscale='linear',  
size='Mail_Total_(tonnes)', color='AustralianPort',  
category='AustralianPort')
```

```
mChart.to_notebook()
```

---

Now I will generate a Motion chart with Passengers\_Total on the x-axis, Freight\_Total(tonnes) on the y-axis and where the color and size of the bubbles is given by the Mail\_Total(tonnes)

```
-----  
filt=(city_reports.AustralianPort=='Melbourne') | (city_reports.AustralianPort=='Perth')  
| (city_reports.AustralianPort=='Sydney') | (city_reports.AustralianPort=='Brisbane')  
city_rep=city_reports[filt]  
#select rows where the AustralianPort is either Melbourne, Perth, Sydney or  
Brisbane
```

```
-----  
city_class=city_rep.groupby(['AustralianPort','Year'])['Passengers_Total','Freight_Tot  
al_(tonnes)','Mail_Total_(tonnes)']  
sample_data=city_class.sum()  
#finding sum for passengers_total,freight_total and mail_total wrt AustralianPort and  
Year
```

```
-----  
from motionchart.motionchart import MotionChart  
import pandas as pd
```

```
-----  
%%html  
<style>  
.output_wrapper, .output {  
    height:auto !important;  
    max-height:1000px; /* your desired max-height here */  
}  
.output_scroll {  
    box-shadow:none !important;  
    webkit-box-shadow:none !important;  
}  
</style>
```

#the following html block code is to ensure that the entire motion chart can be seen nicely in the output cell



---

```
mChart = MotionChart(df = sample_data)
mChart.to_notebook()
#outputting the motion chart to see if the options and parameters in the motionchart
are what we need
```

---

```
mChart = MotionChart(df = sample_data, key='Year', x='Passengers_Total',
y='Freight_Total_(tonnes)', xscale='linear', yscale='linear',
                        size='Mail_Total_(tonnes)', color='Mail_Total_(tonnes)',
category='AustralianPort')
```

```
mChart.to_notebook()
#we are resetting the options and parameters in the motionchart here
#this is the final motionchart we will be using to answer questions related to Airport
Traffic
```

---