BT1101: Case Assignment 1

Ow Wen Hui Shirley

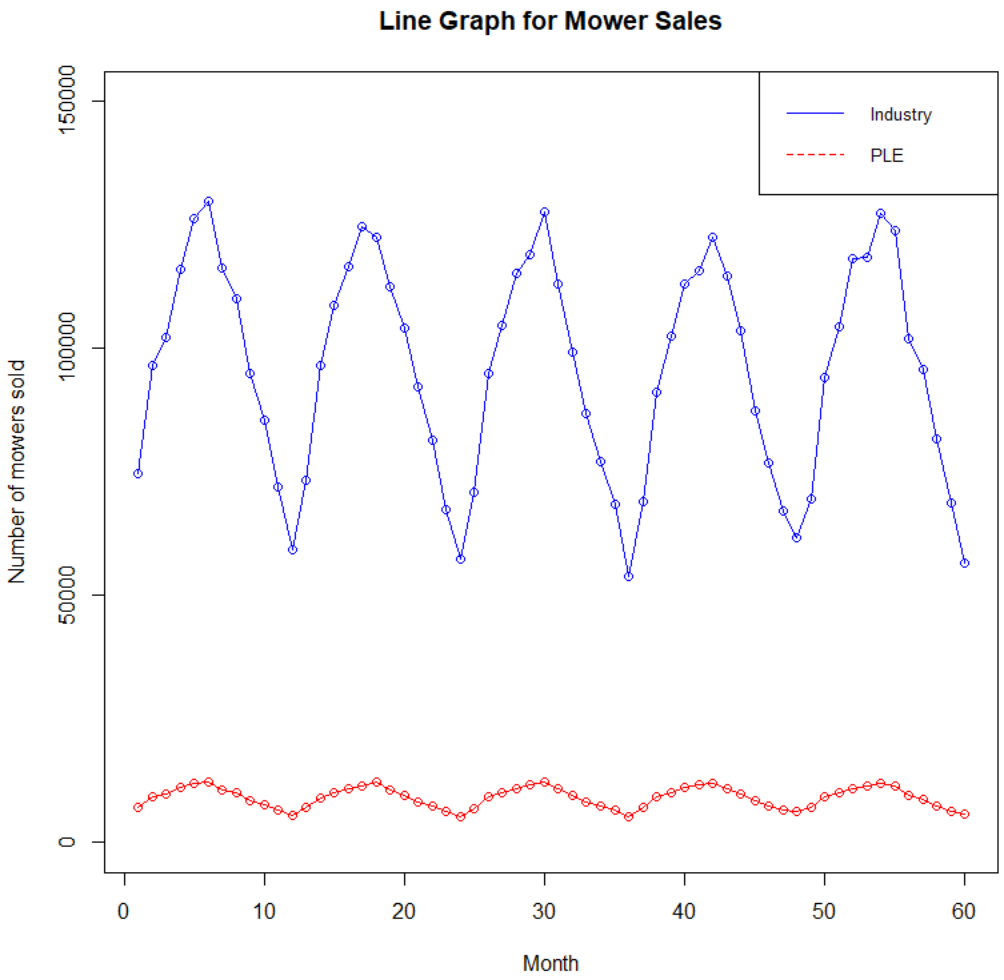
Maegan Leong

1) Understanding the data

2i) How sales of mowers and tractors compare with industry totals and how strongly monthly product sales are correlated with industry sales.

To compare the sales of mowers with industry totals, we have plotted 2 lines – one representing the total sales of PLE mowers in the world over time and another representing the total industry sales of mowers in the world over time – on a line graph.

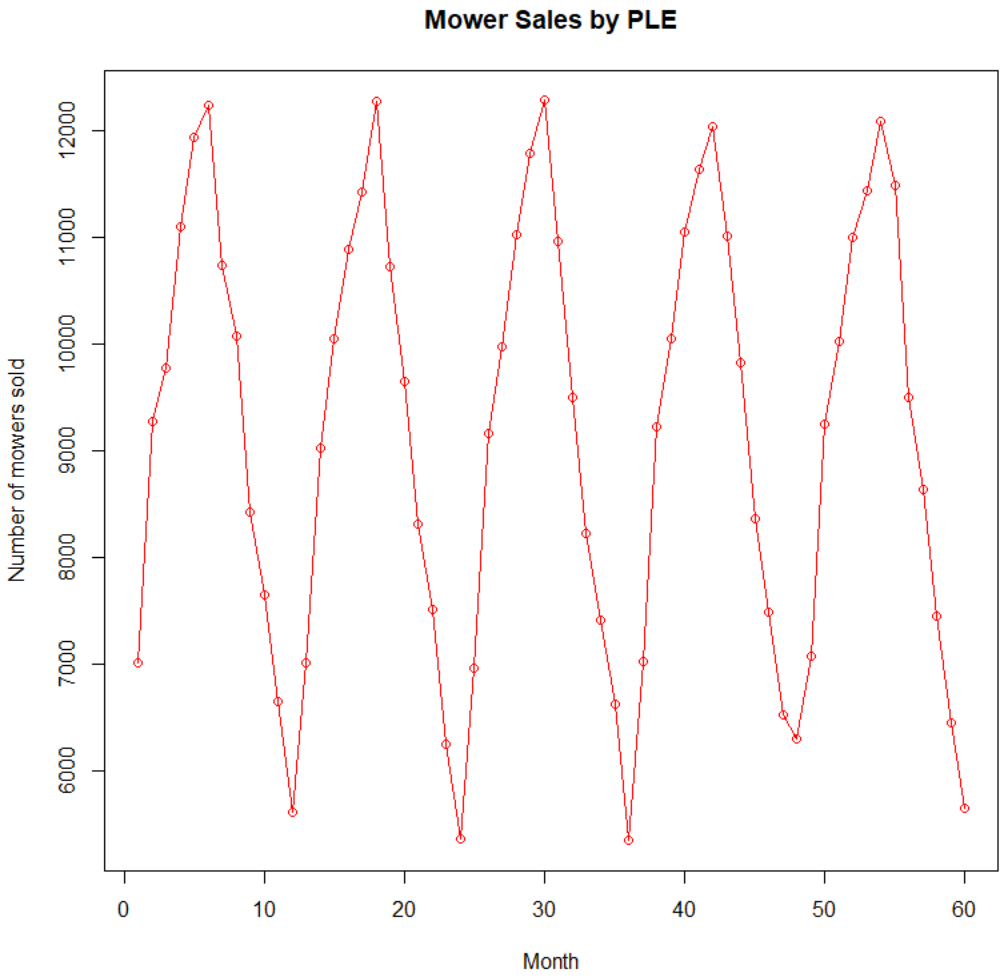
The x-axis of the line graph below is represented by Month 1 to Month 60, assuming that Month 1 starts from January 2010 and Month 60 ends in December 2014. The y-axis of the line graph shows the number of mowers sold.



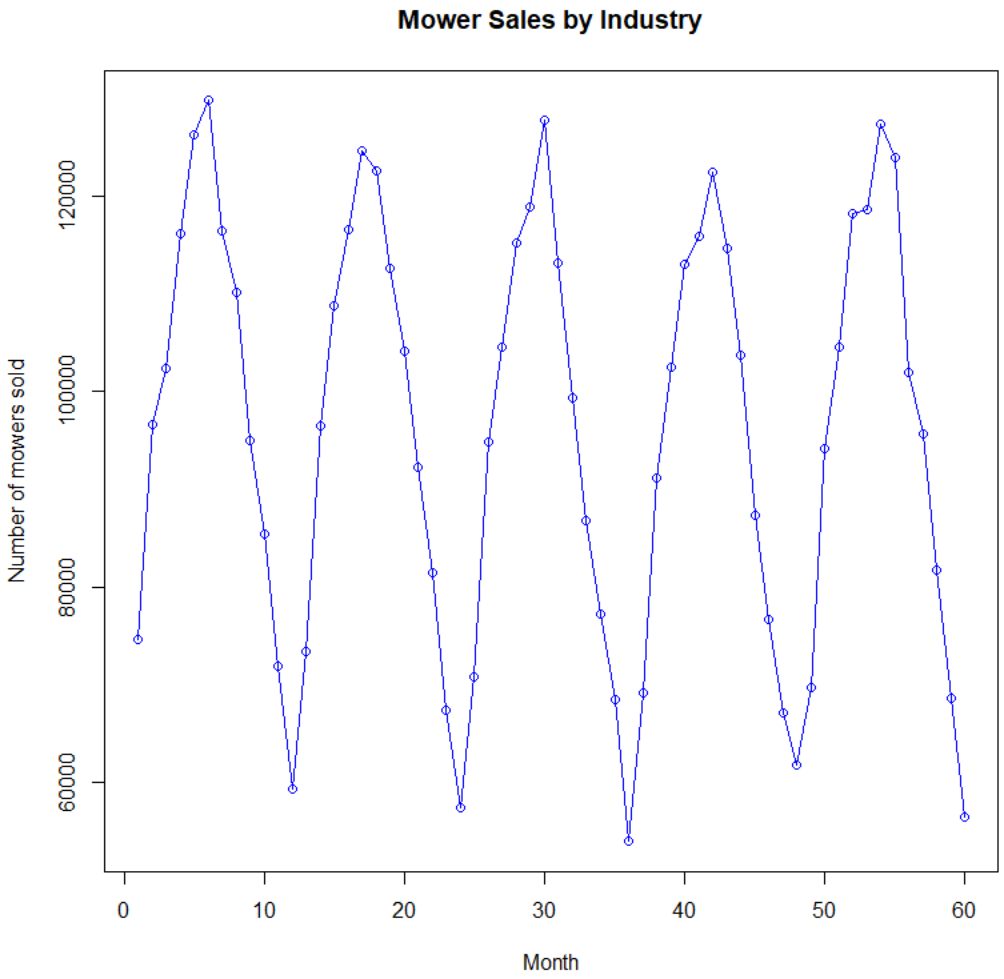
Graph 1: The line chart above compares the total world sales of mowers by the industry and the total world sales of mowers by the firm PLE over the course of 5 years, from the start of 2010 to the end of 2014.

Inference from Graph 1:

- Industry sales of mowers are seen here to be much higher than those of PLE, suggesting that PLE does not dominate the world market of mowers (market share is less than 50%). There also appears to be similar periodic fluctuations over the 5 years between PLE's sales of mowers and the industry's total sales of mowers.



Graph 2: A zoomed-in line graph showing the total world sales of PLE mowers over the course of 5 years.



Graph 3: A zoomed-in line graph showing the total world sales of mowers in the industry over the course of 5 years.

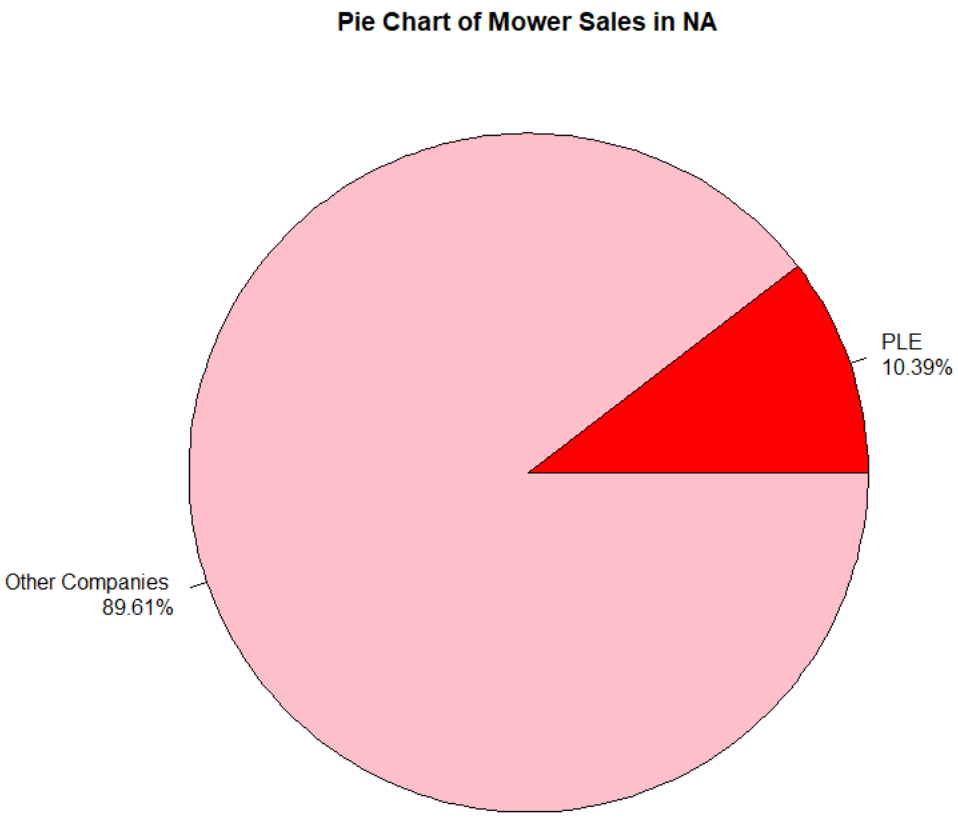
Inference from Graphs 2 and 3:

- It can be observed that there are very similar fluctuation patterns in both line graphs over the course of the 60 months, with PLE's mower sales taking up around 1/10 of the industry's mower sales.

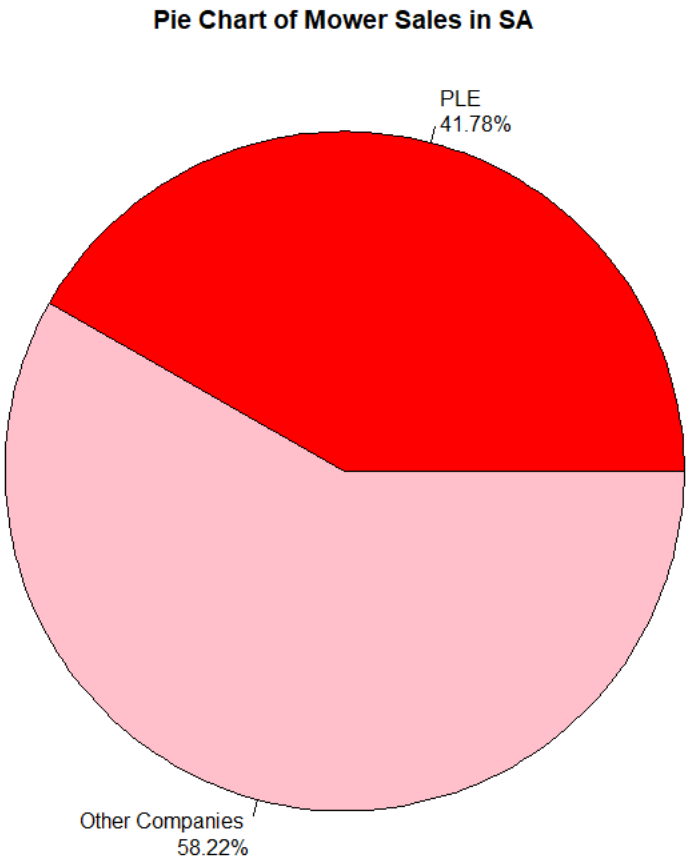
Next, to analyse PLE's market share in the mowers markets in the different regions, we drew up pie charts to show the proportions of their market shares in the markets of each of the regions.

The following pie charts show the percentage of total mower sales in each region in the 5 years. China is not represented because there is no data on China's industry sales for mowers. The entire pie represents the total number of tractor units sold in the industry in the 5 years, and hence these pie charts show an overall market share (in %) PLE mowers take up in each of the regions, in relative to the market share taken up by all other firms in the mower-producing industry in the region.

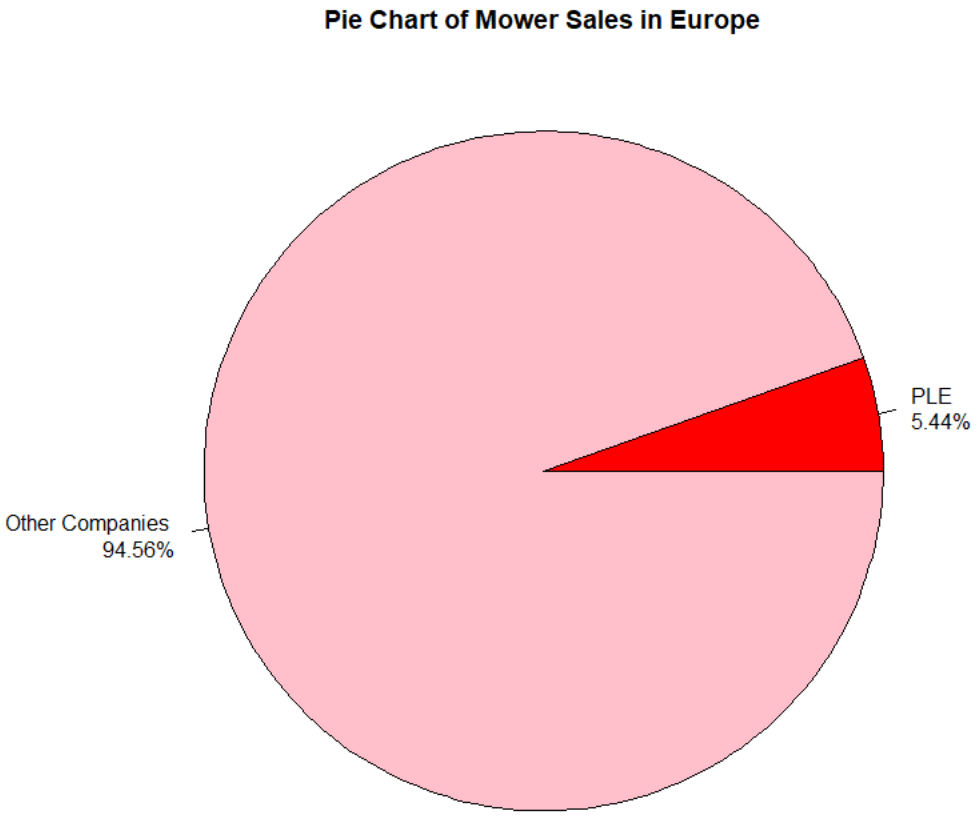
These percentages were obtained by taking the total number of mower units sold by PLE in the 5 years in the specific region divided by the total number of mower units sold by the industry in the 5 years in the specific region, multiplied by 100.



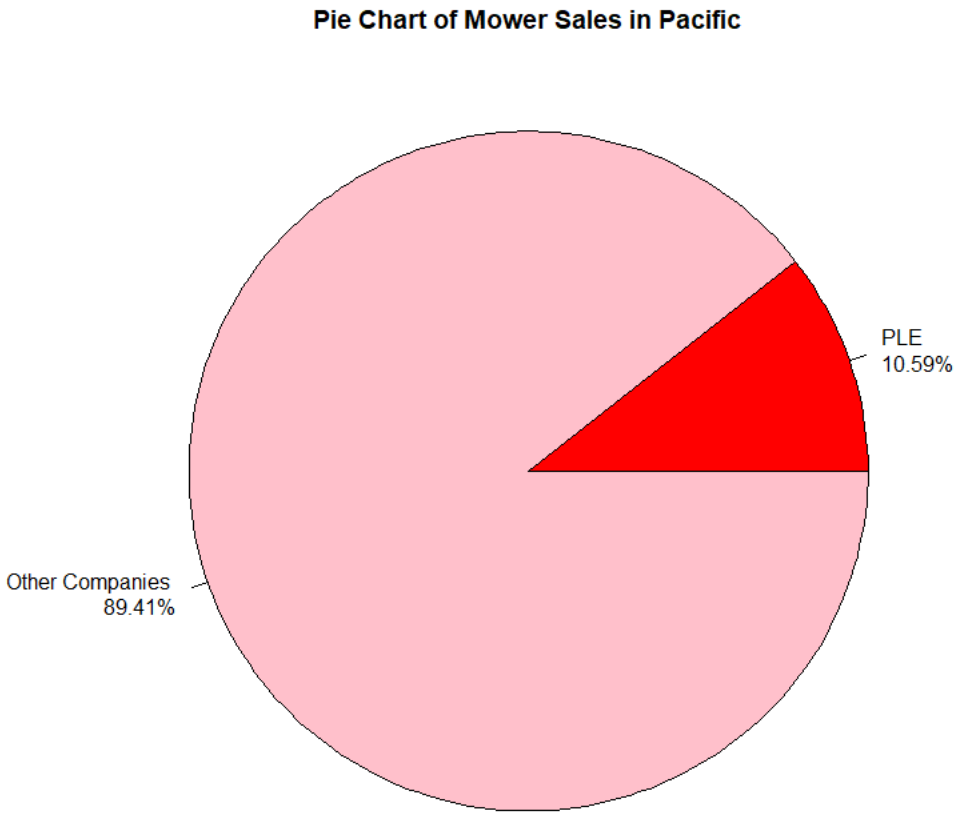
Graph 4: A pie chart of mower sales in North America, representing overall PLE's market share in the mower industry in North America.



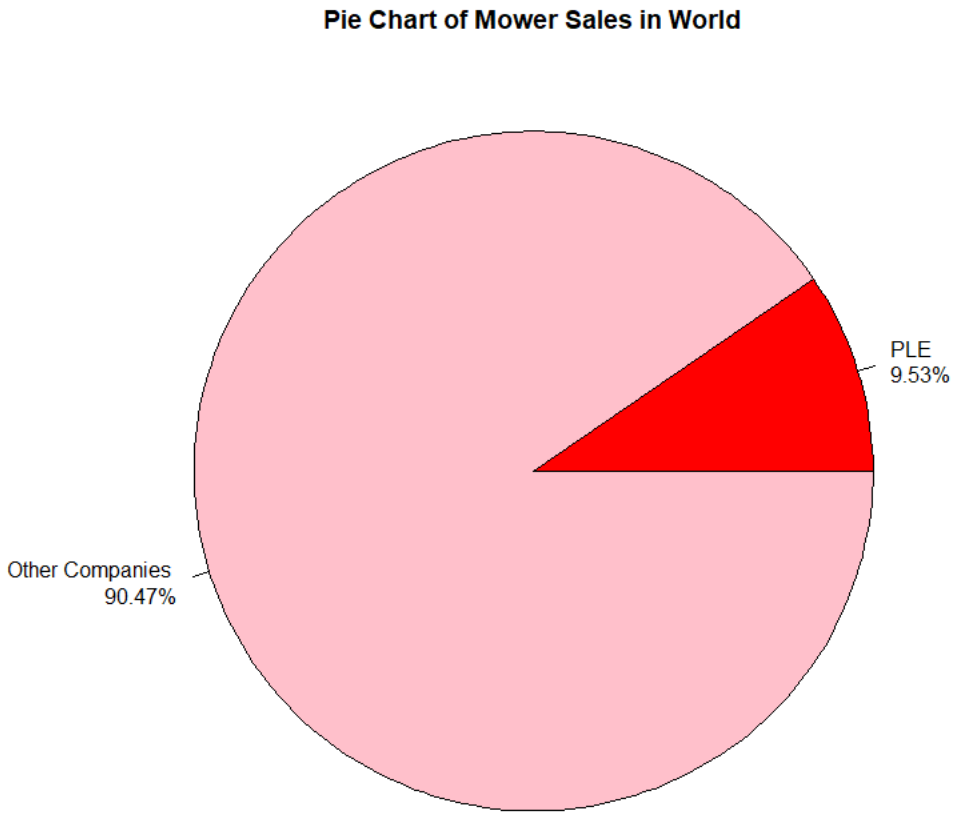
Graph 5: A pie chart of mower sales in South America, representing overall PLE's market share in the mower industry in South America.



Graph 6: A pie chart of mower sales in Europe, representing PLE's overall market share in the mower industry in Europe.



Graph 7: A pie chart of mower sales in Pacific, representing PLE's overall market share in the mower industry in Pacific.



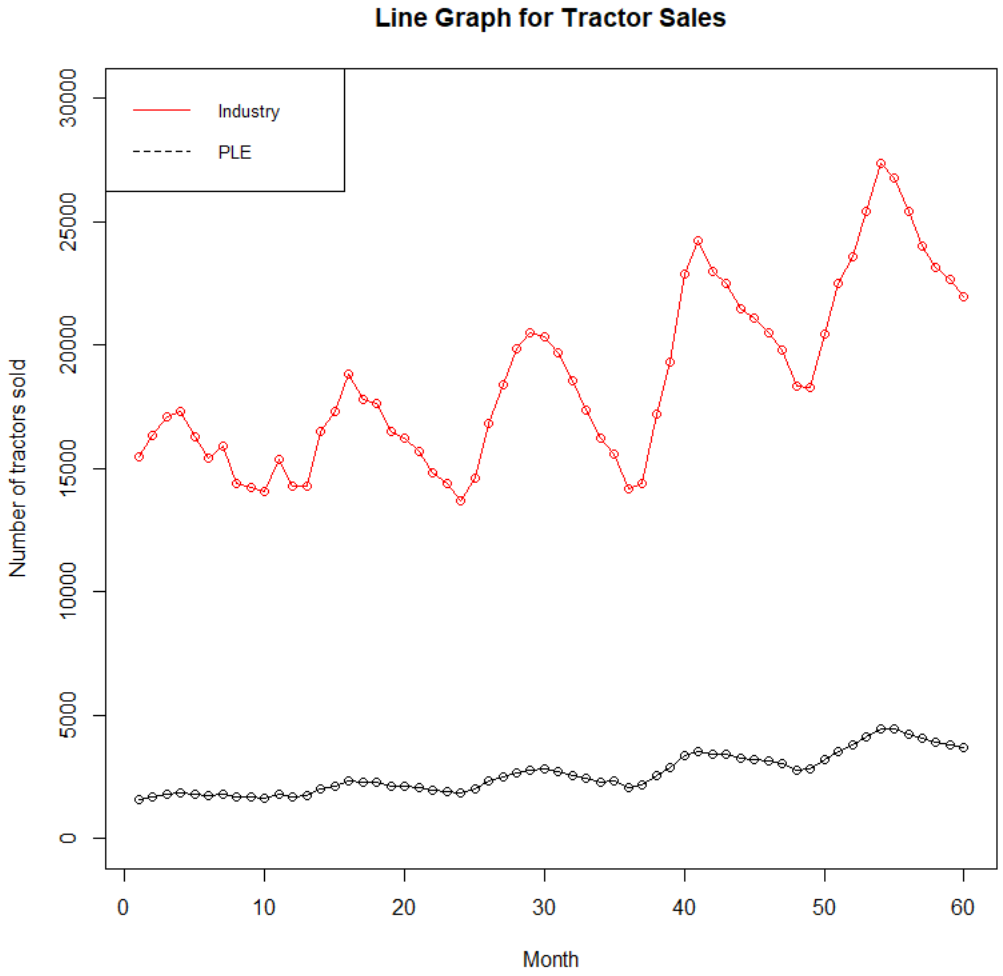
Graph 8: A pie chart of mower sales in the world, representing PLE's overall market share in the mower industry in the world.

Conclusion:

- Based on the line graphs of the world sales of mowers, PLE mowers take up around one-tenth of the total number of mower units sold in the world, and PLE follows similar cyclical sales patterns as those of the industry.
- PLE's overall market share in the world mower industry is 9.53%.
- Based on sales of mowers in individual regions, PLE has the highest overall market share in South America (41.78%) and the lowest overall market share in Europe (5.44%). PLE's overall market share in North America is 10.39% and its overall market share in Pacific is 10.59%.

To compare the sales of tractors with industry totals, we used a similar method as when we were comparing the sales of mowers. We have plotted 2 lines – one representing the total sales of PLE tractors in the world over time and another representing the total industry sales of tractors in the world over time – on a line graph.

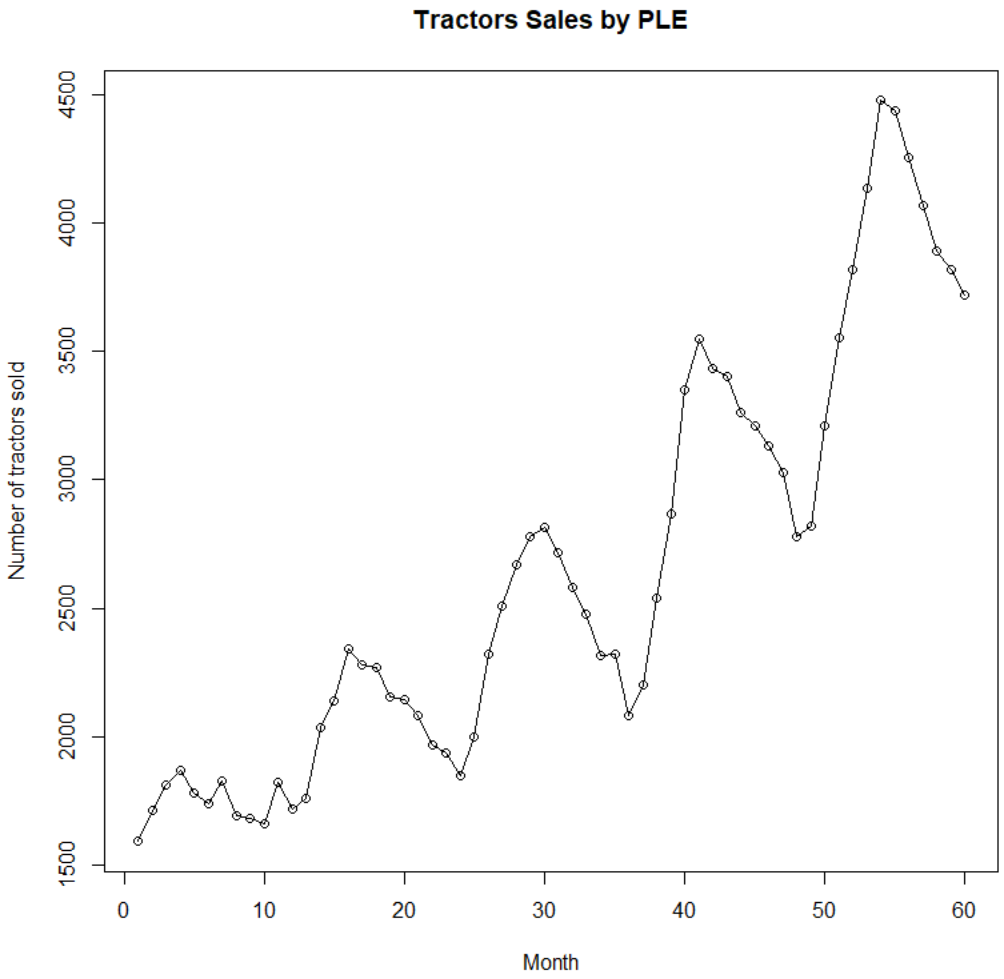
The x-axis of the line graph below is represented by Month 1 to Month 60, assuming that Month 1 starts from January 2010 and Month 60 ends in December 2014. The y-axis of the line graph shows the number of tractors sold.



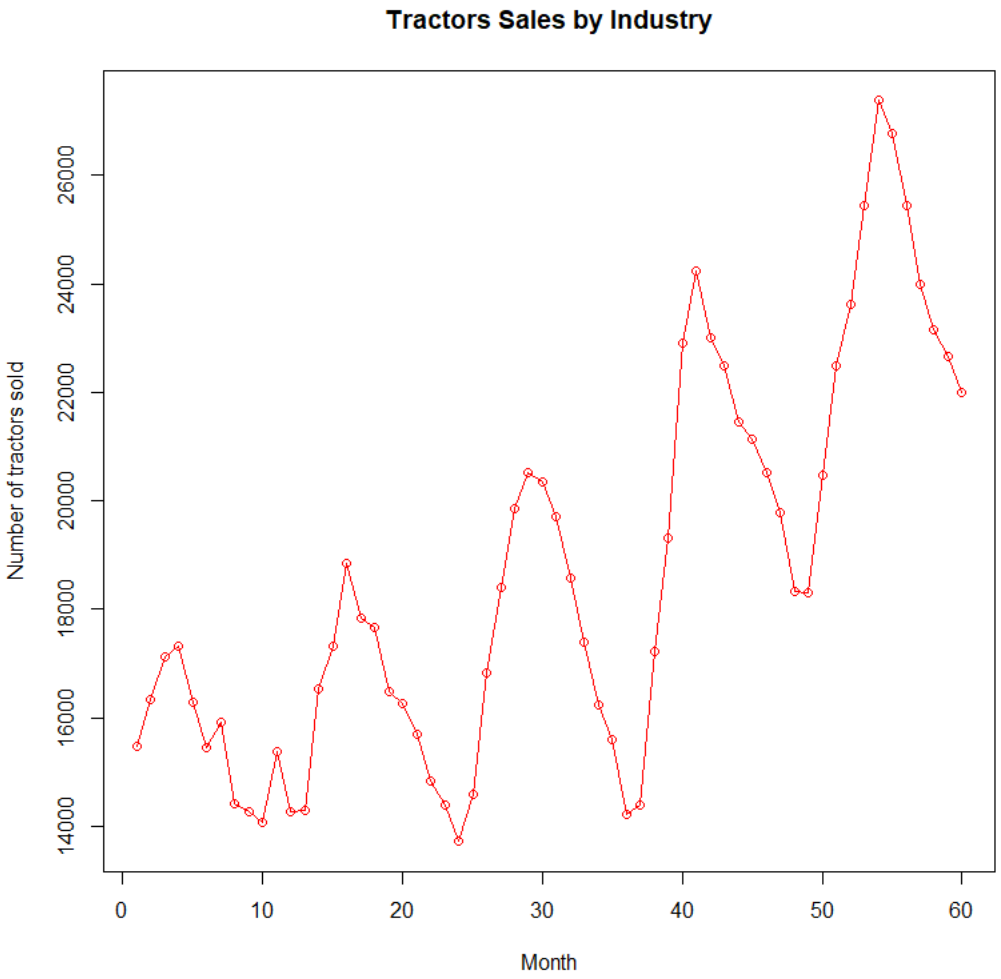
Graph 9: The line chart above compares the total world sales of tractors by the industry and the total world sales of tractors by the firm PLE over the course of 5 years, from the start of 2010 to the end of 2014.

Inference from Graph 9:

- Industry sales of tractors are seen here to be much higher than those of PLE, suggesting that PLE does not dominate the world market of tractors (market share is less than 50%). There also appears to be similar periodic fluctuations over the 5 years between PLE's sales of tractors and the industry's total sales of tractors.



Graph 10: A zoomed-in line graph showing the total world sales of PLE tractors over the course of 5 years.



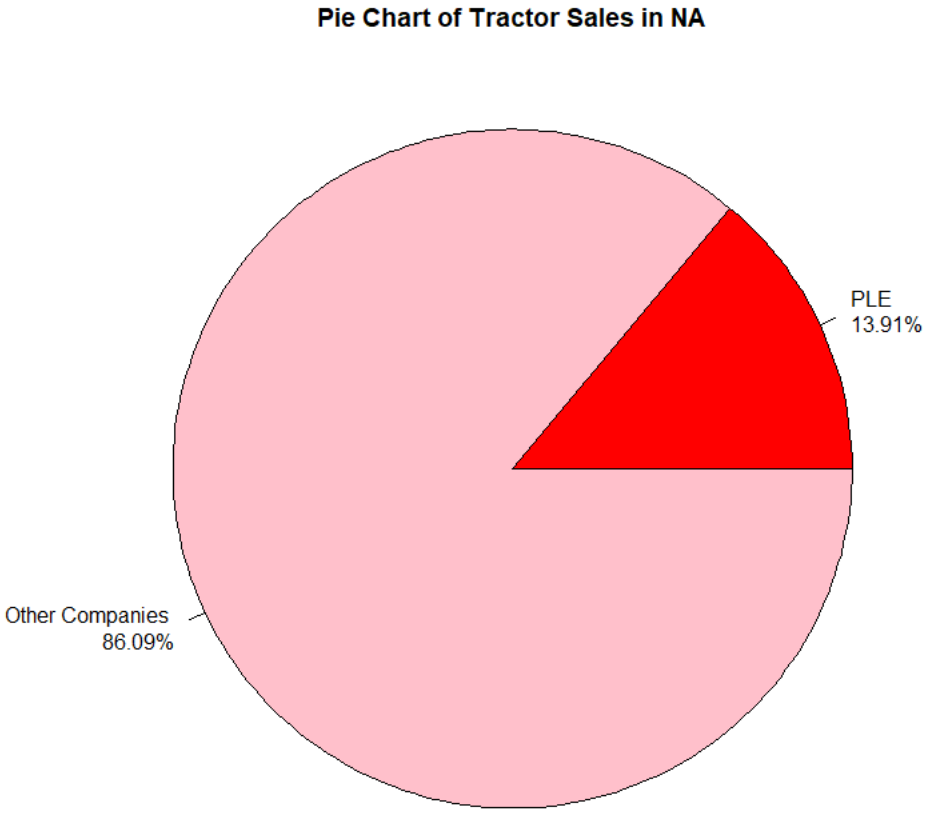
Graph 11: A zoomed-in line graph showing the total world sales of industry tractors over the course of 5 years.

Inference for Graphs 10-11:

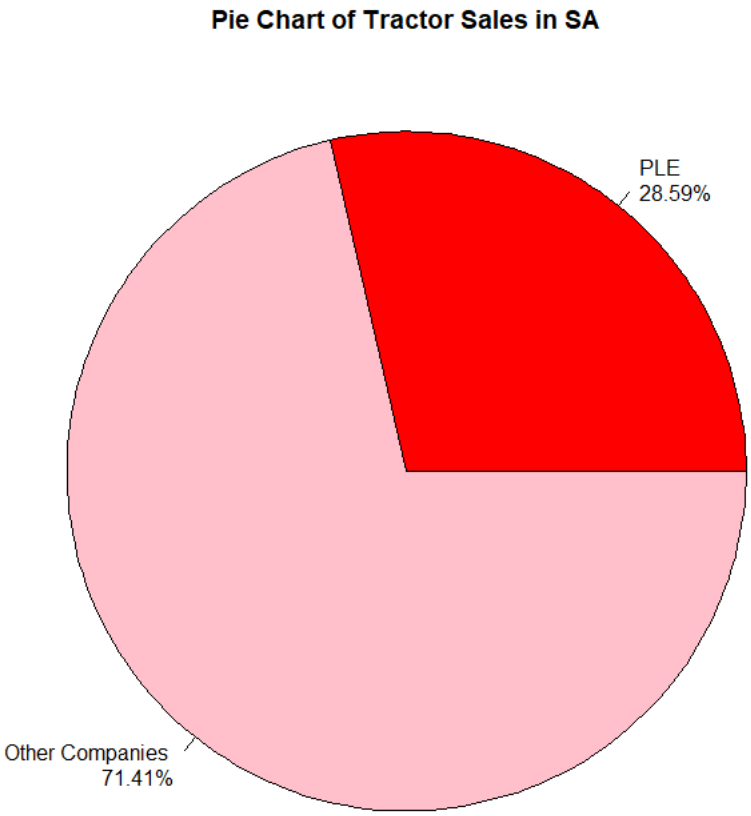
- It can be observed that there are some similar fluctuation patterns in both line graphs over the course of the 60 months, and a common upward trend of number of tractors sold over the years.
- In Month 1, PLE's tractor sales was around 1/10 of industry's tractor sales.
- In some of the months, PLE's tractor sales do not follow the fluctuation patterns of the industry's tractor sales closely, implying changes in market share in the 5 years. For example, in the 36th month, the tractor industry's sales were low at 14,207, but tractor sales by PLE was higher than the expected 1,400 (assuming PLE's tractor sales taking up around 1/10 of the industry's tractor sales). PLE sold 2,080 tractors in the 36th month, which makes PLE's market share in the world around 14.6%. Hence, PLE's market share in the world market of tractors fluctuates, unlike PLE's market share in the world market of mowers, which is more constant as PLE's mower sales follow the fluctuation patterns of the industry's mower sales closely.

Next, to analyse PLE's market share in the tractors markets in the different regions, we drew up pie charts to show the proportions of their market shares in the markets of each of the regions.

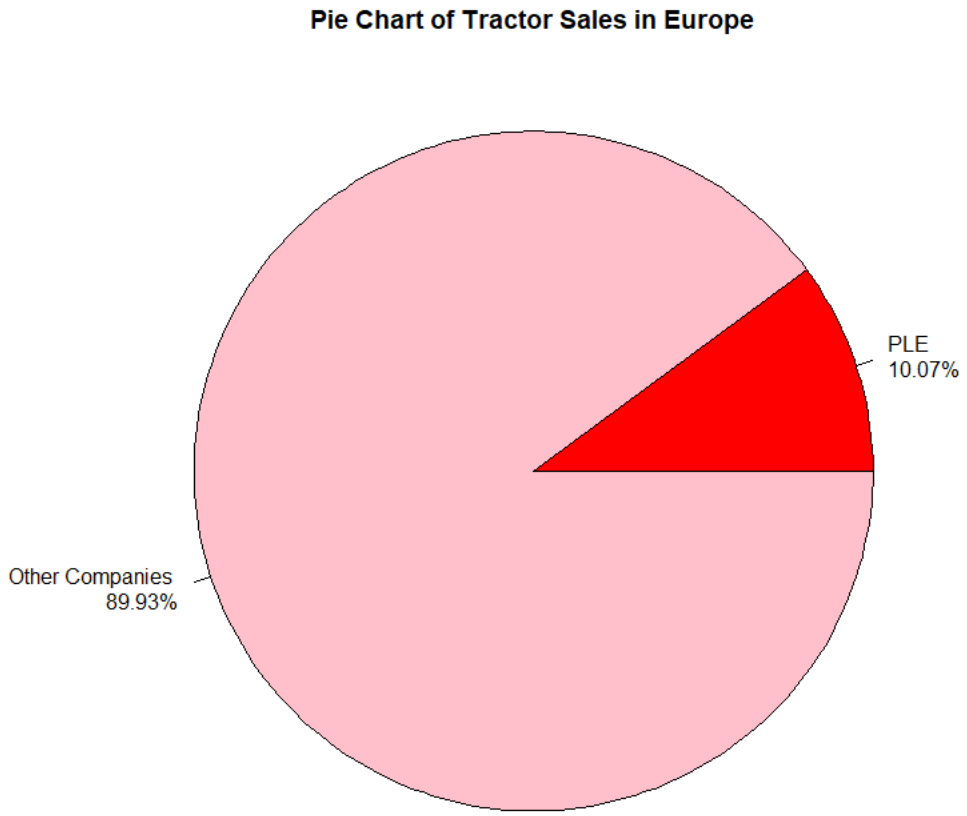
The following pie charts show the percentage of tractor sales in each region throughout the 5 years. The entire pie represents the total number of tractor units sold in the industry in the 5 years, and hence these pie charts show an overall market share PLE tractors take up in each of the regions, in relative to the market share taken up by all other firms in the tractor-producing industry.



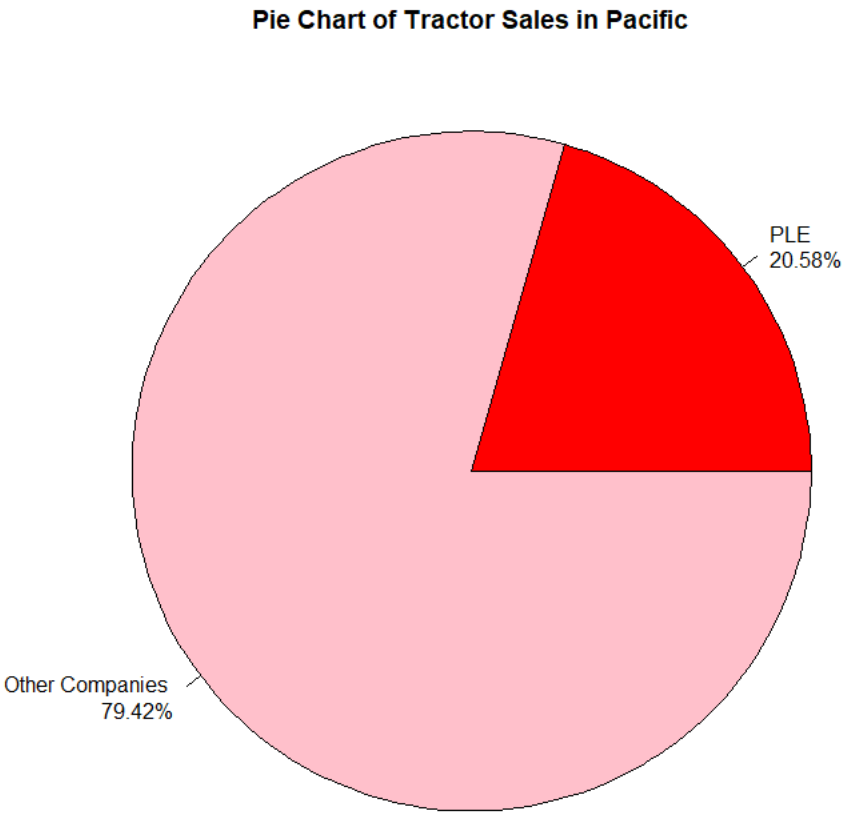
Graph 12: A pie chart of tractor sales in North America, representing PLE's market share in the tractor industry in North America.



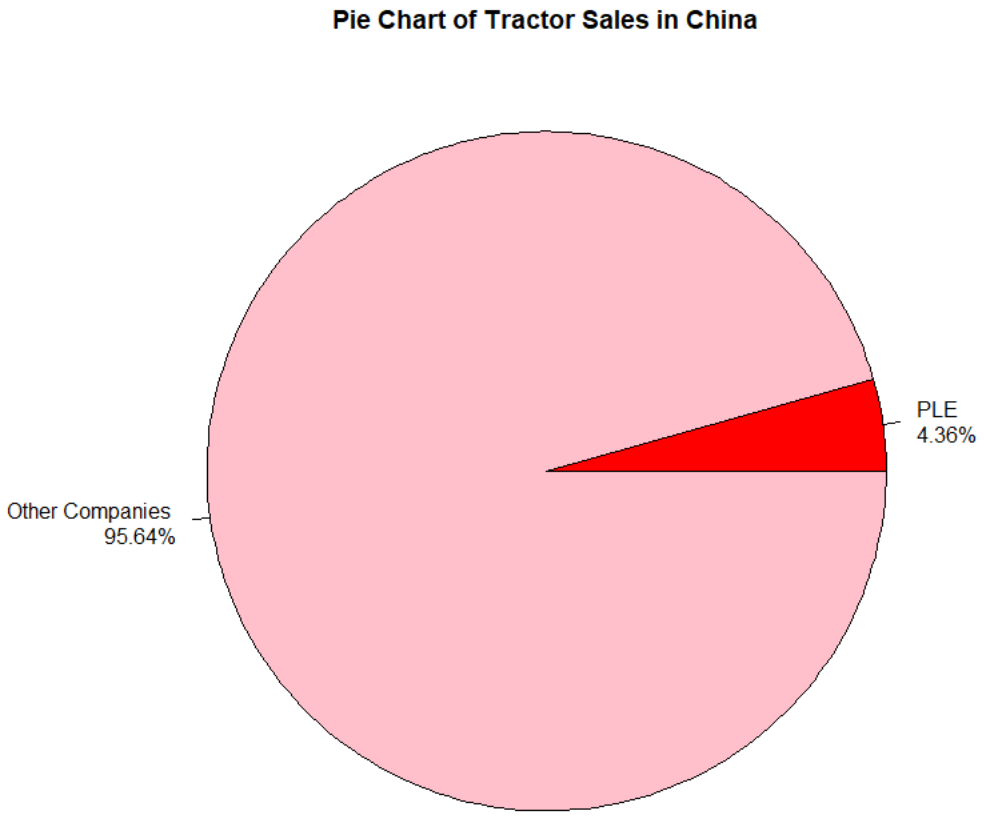
Graph 13: A pie chart of tractor sales in South America, representing PLE's market share in the tractor industry in South America.



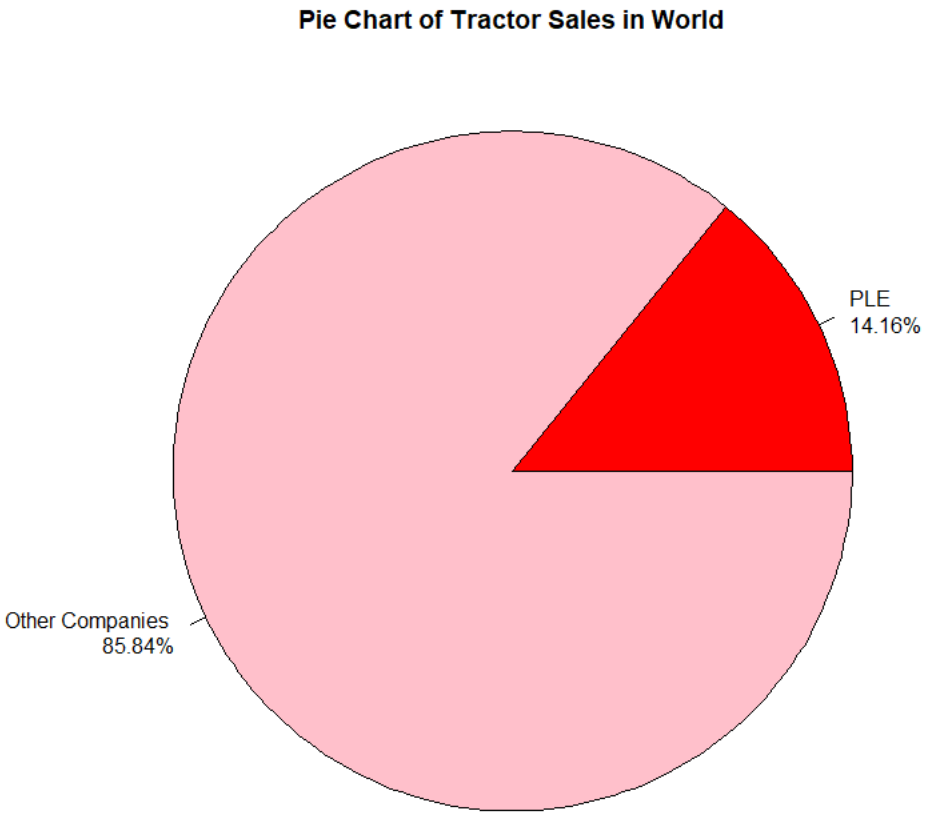
Graph 14: A pie chart of tractor sales in Europe, representing PLE's market share in the tractor industry in Europe.



Graph 15: A pie chart of tractor sales in Pacific, representing PLE's market share in the tractor industry in Pacific.



Graph 16: A pie chart of tractor sales in China, representing PLE's market share in the tractor industry in China.

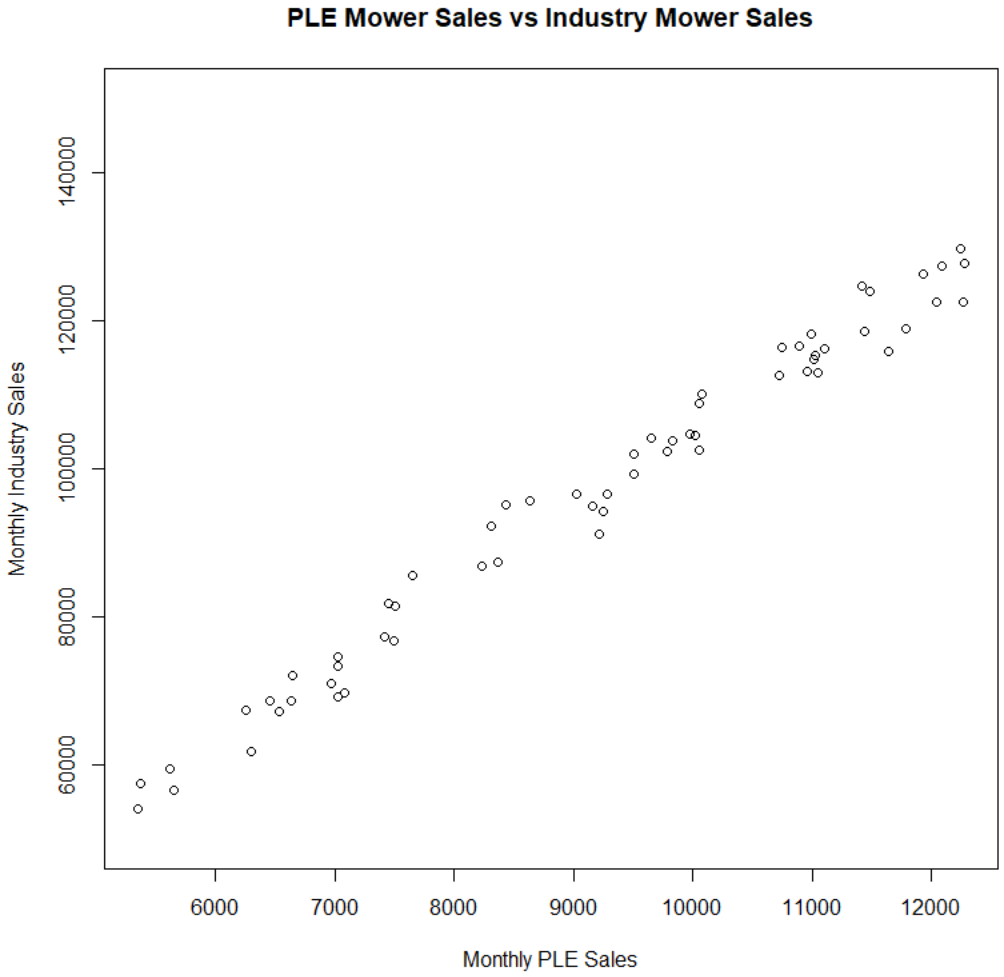


Graph 17: A pie chart of tractor sales in the world, representing PLE's market share in the tractor industry in the world.

Conclusion:

- PLE's market share in the world tractors market fluctuates over the 5 years (between 10.2% and 16.9%), because PLE's world tractor sales do not follow the fluctuation patterns of the industry's world tractor sales closely.
- PLE's overall market share in the world tractors market is 14.16%.
- Based on sales of tractors in individual regions, PLE has the highest overall market share in South America's market of tractors (28.59%), and the lowest overall market share in China (4.36%). This is understandable because PLE has only just opened the tractors market in China 3 years ago, and hence since PLE has just entered China's market, its market share in China is still low. As stated in the excerpt in the extract for this assignment, "PLE's sales include… growing South American market", South America's market is growing, and that is probably why PLE's largest market shares for both mowers and tractors markets are in South America.

To examine how strongly monthly product sales are correlated with industry sales, scatterplots and correlation coefficients are used.



Graph 18: A scatterplot of monthly Industry mower sales in the world against monthly PLE mower sales in the world.

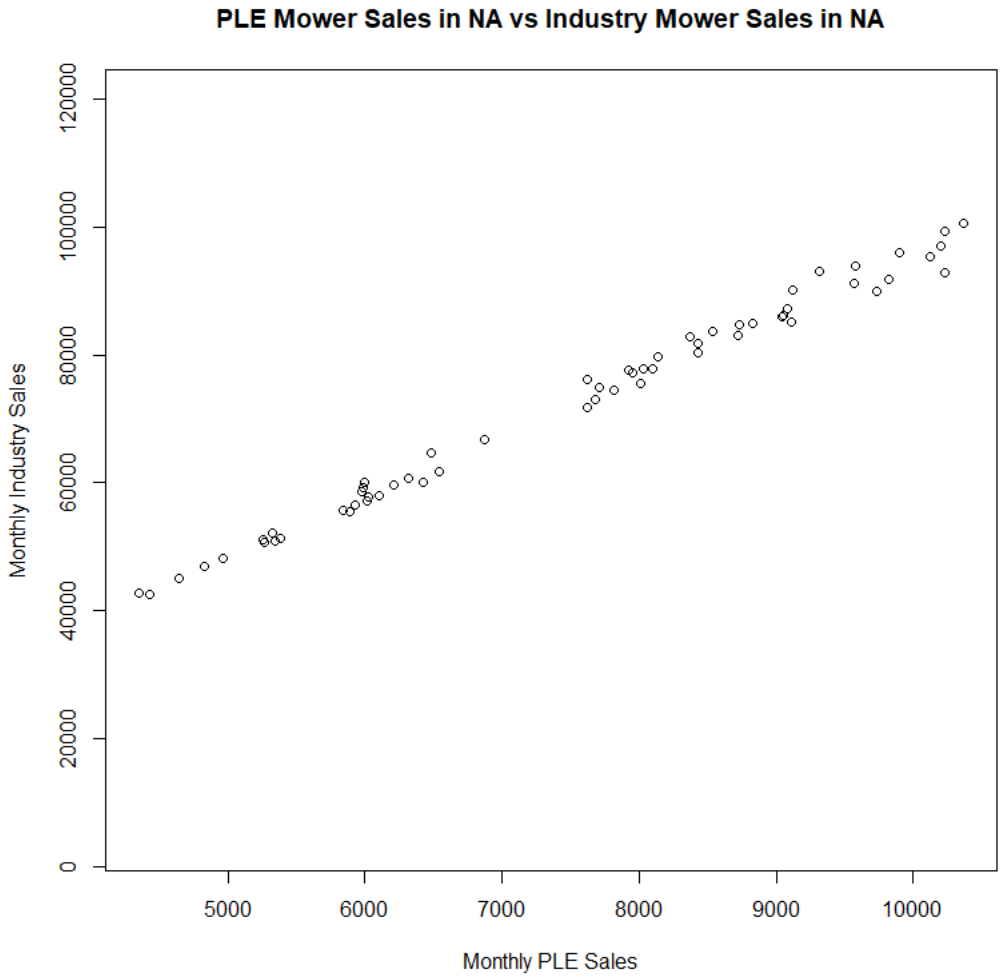
Correlation Coefficient: 0.9904799

Inference for Graph 18 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Mower Sales in the world and Monthly Industry Mower Sales in the world. This means that as Monthly PLE Mower Sales increases, Monthly Industry Mower Sales tends to increase at a constant rate.

Similarly, we can draw scatterplots of monthly Industry mower sales in each of the different regions against monthly PLE mower sales in each of the different regions and calculate the

correlation coefficients to examine the strength of correlation between the monthly Industry mower sales and monthly PLE mower sales in each region.

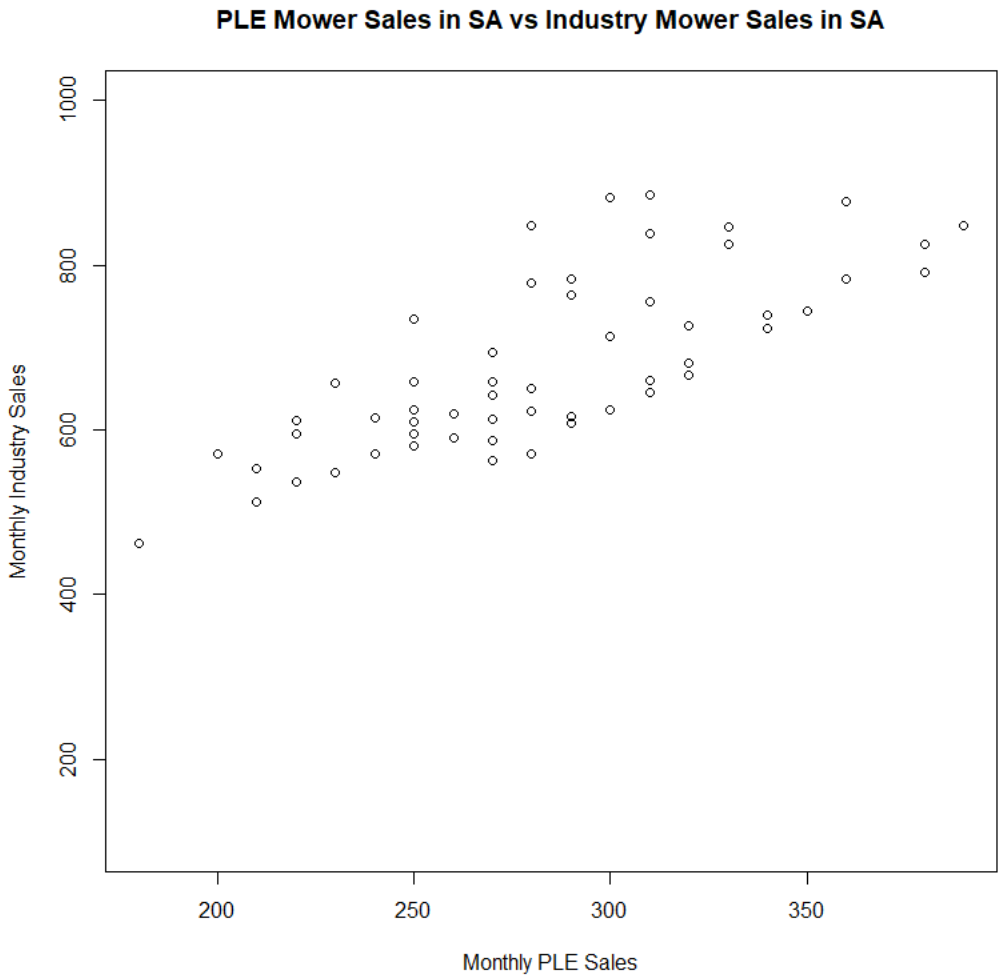


Graph 19: A scatterplot of monthly Industry mower sales in North America against monthly PLE mower sales in North America.

Correlation Coefficient: 0.9957983

Inference for Graph 19 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Mower Sales in North America and Monthly Industry Mower Sales in North America. This means that as Monthly PLE Mower Sales increases, Monthly Industry Mower Sales tends to increase at a constant rate.

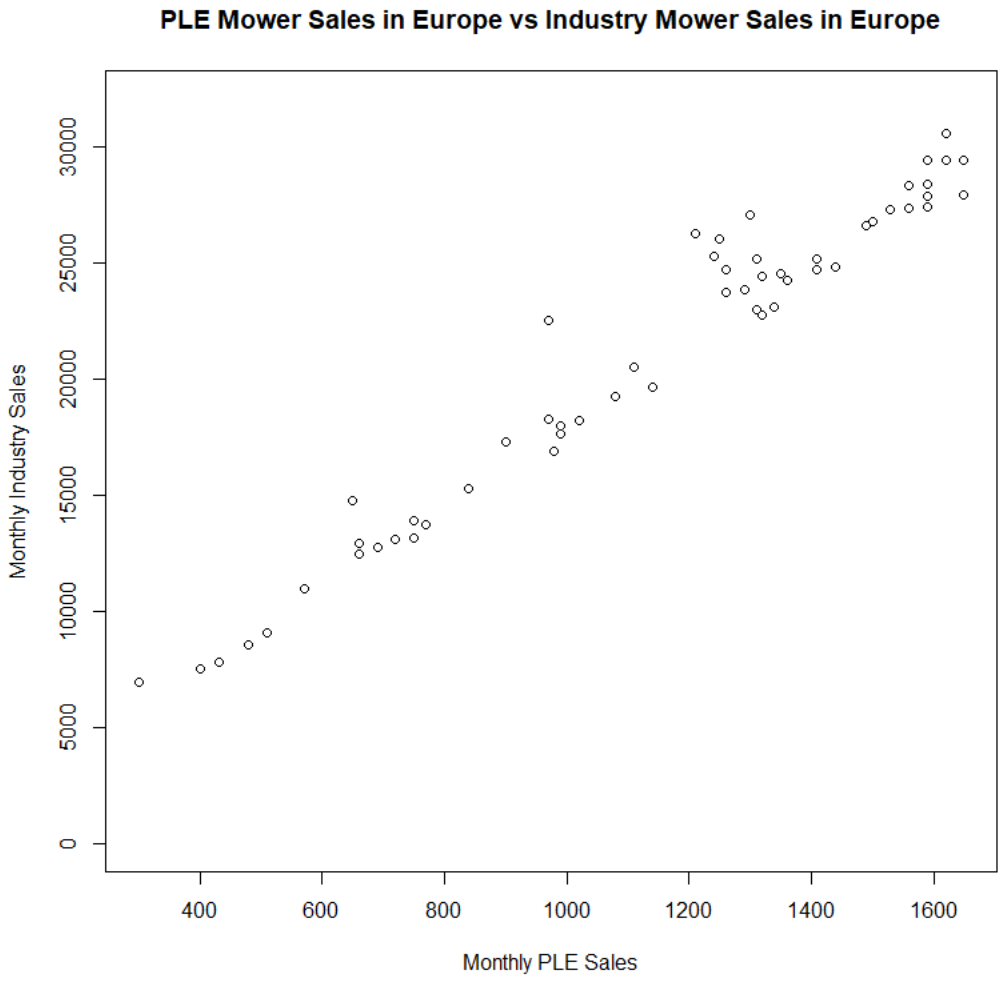


Graph 20: A scatterplot of monthly Industry mower sales in South America against monthly PLE mower sales in South America.

Correlation Coefficient: 0.7571162

Inference for Graph 20 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is close to +1, there could be a moderate positive linear correlation between Monthly PLE Mower Sales in South America and Monthly Industry Mower Sales in South America. This means that as Monthly PLE Mower Sales increases, Monthly Industry Mower Sales tends to increase at a constant rate.

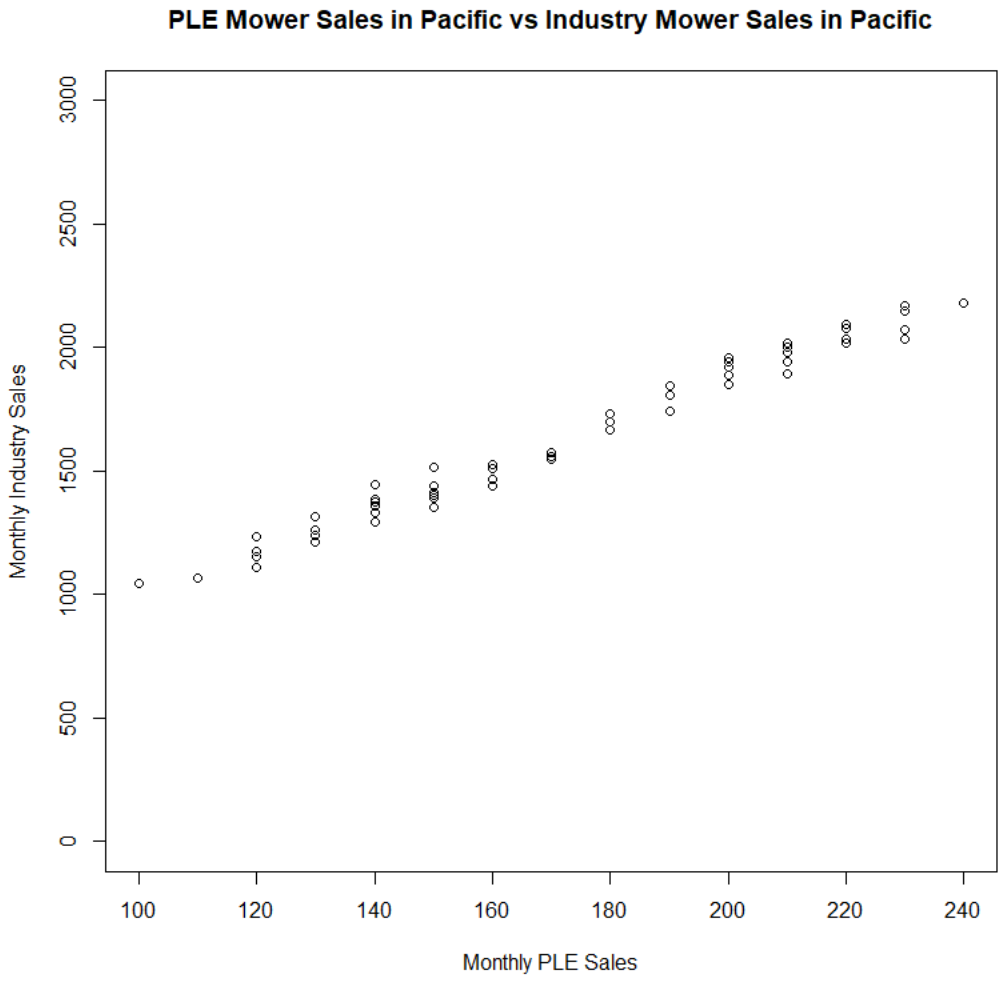


Graph 21: A scatterplot of monthly Industry mower sales in Europe against monthly PLE mower sales in Europe.

Correlation Coefficient: 0.9792085

Inference for Graph 21 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Mower Sales in Europe and Monthly Industry Mower Sales in Europe. This means that as Monthly PLE Mower Sales increases, Monthly Industry Mower Sales tends to increase at a constant rate.



Graph 22: A scatterplot of monthly Industry mower sales in Pacific against monthly PLE mower sales in Pacific.

Correlation Coefficient: 0.9901594

Inference for Graph 22 & Correlation Coefficient:

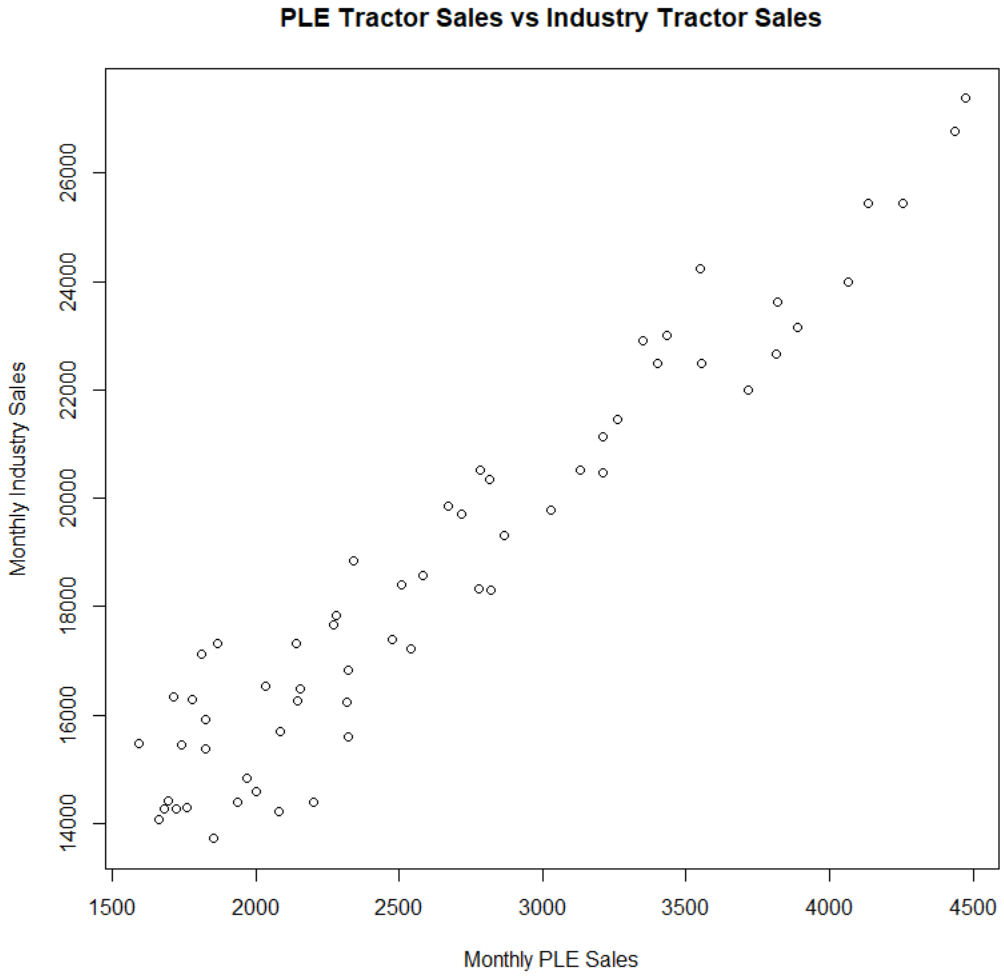
- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Mower Sales in Pacific and Monthly Industry Mower Sales in Pacific. This means that as Monthly PLE Mower Sales increases, Monthly Industry Mower Sales tends to increase at a constant rate.

Conclusion:

- PLE monthly mower sales in the world is strongly linearly correlated to the industry's monthly mower sales in the world.
- PLE monthly mower sales in all of the regions seem to be strongly correlated to the industry monthly mower sales in each of the regions.

- For South America however, the linear correlation between PLE monthly mower sales and industry monthly mower sales was only moderately strong, weaker as compared to those of other regions.

Similarly, for the tractor's market, a similar analysis can be made.



Graph 23: A scatterplot of monthly Industry tractor sales in the world against monthly PLE tractor sales in world.

Correlation Coefficient: 0.9602768

Inference for Graph 23 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in the world and Monthly Industry Tractor Sales in the world. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.

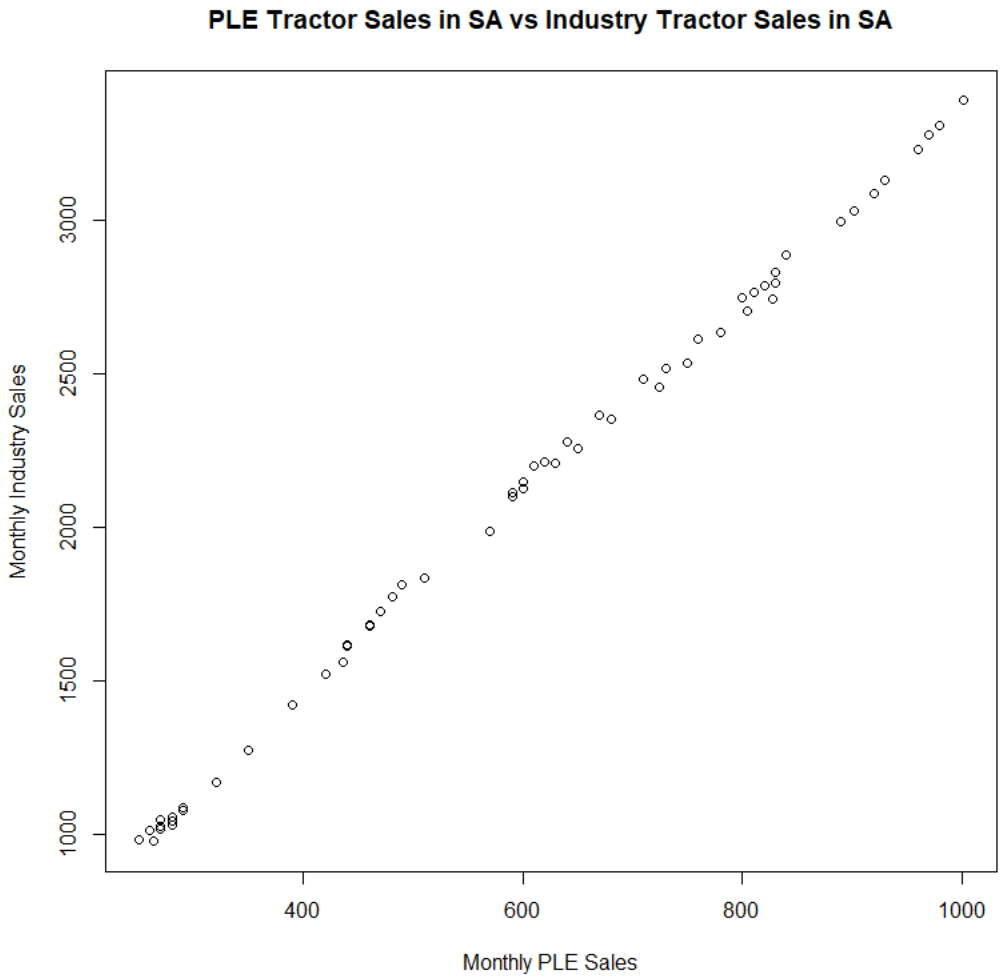
To examine the strength of correlation in the sales of each region, we can draw scatterplots of monthly Industry tractor sales in each of the different regions against monthly PLE tractor sales in each of the different regions and calculate the correlation coefficients.

Graph 24: A scatterplot of monthly Industry tractor sales in North America against monthly PLE tractor sales in North America.

Correlation Coefficient: 0.8934245

Inference for Graph 24 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in North America and Monthly Industry Tractor Sales in North America. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.

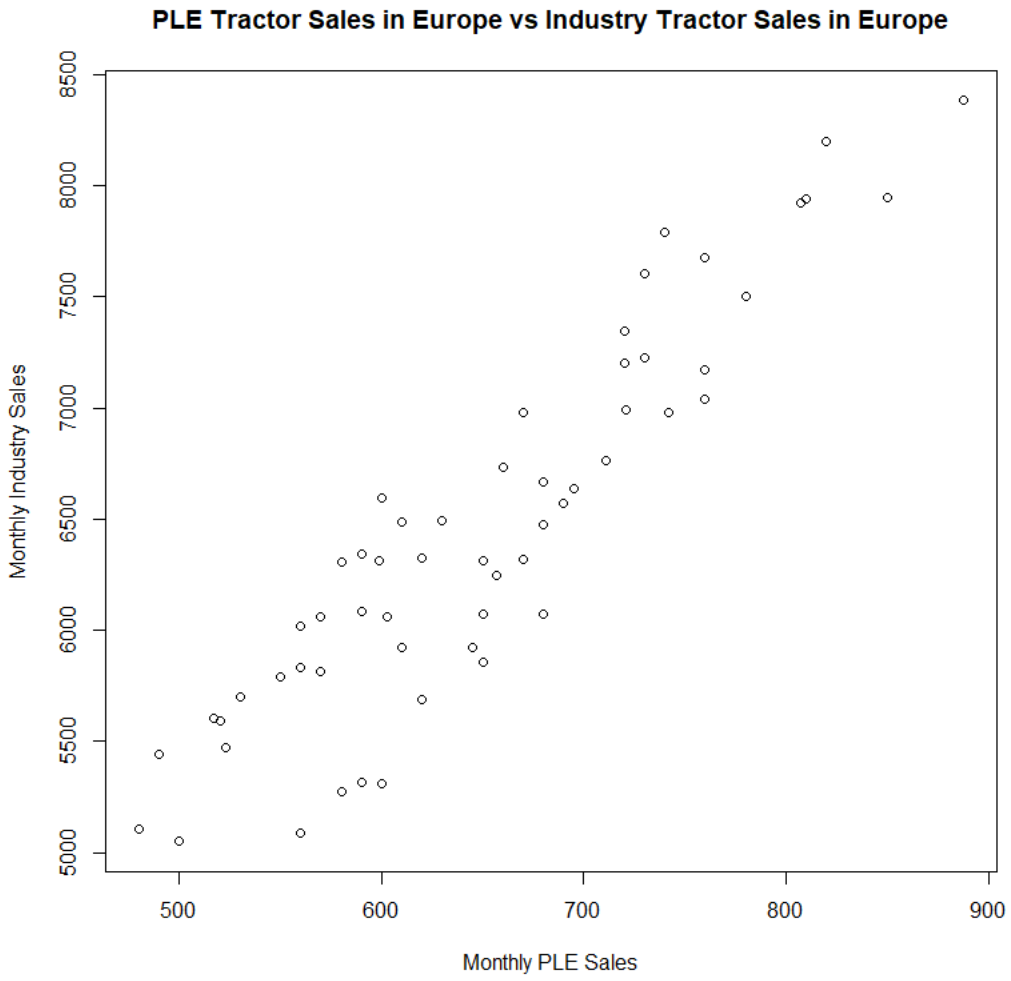


Graph 25: A scatterplot of monthly Industry tractor sales in South America against monthly PLE tractor sales in South America.

Correlation Coefficient: 0.999017

Inference for Graph 25 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in South America and Monthly Industry Tractor Sales in South America. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.

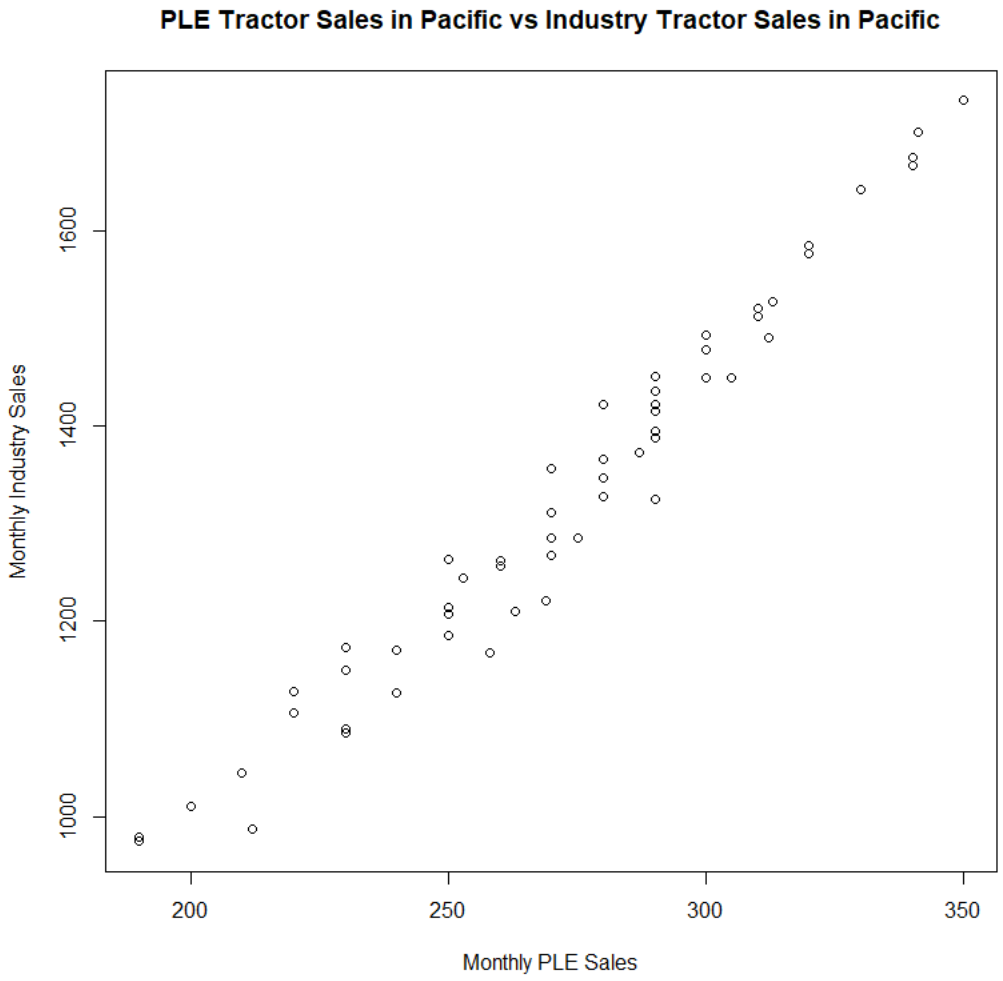


Graph 26: A scatterplot of monthly Industry tractor sales in Europe against monthly PLE tractor sales in Europe.

Correlation Coefficient: 0.9186482

Inference for Graph 26 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in Europe and Monthly Industry Tractor Sales in Europe. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.

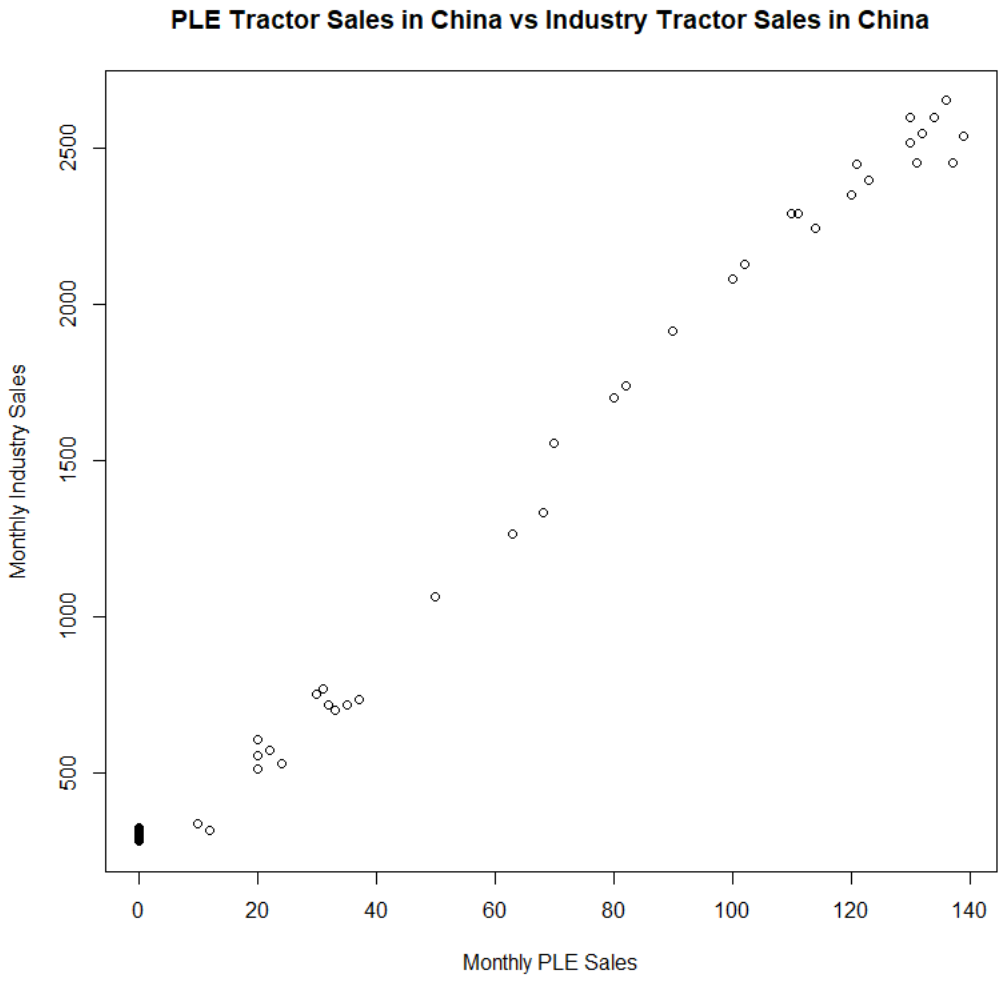


Graph 27: A scatterplot of monthly Industry tractor sales in Pacific against monthly PLE tractor sales in Pacific.

Correlation Coefficient: 0.9803292

Inference for Graph 27 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in Pacific and Monthly Industry Tractor Sales in Pacific. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.



Graph 28: A scatterplot of monthly Industry tractor sales in China against monthly PLE tractor sales in China.

Correlation Coefficient: 0.9962615

Inference for Graph 28 & Correlation Coefficient:

- Since points lie close to a straight line with positive gradient, and the value of r is very close to +1, there could be a strong positive linear correlation between Monthly PLE Tractor Sales in China and Monthly Industry Tractor Sales in China. This means that as Monthly PLE Tractor Sales increases, Monthly Industry Tractor Sales tends to increase at a constant rate.

Conclusion:

- PLE monthly tractor sales in the world is strongly linearly correlated to the industry's monthly tractor sales in the world.
- PLE monthly tractor sales in all of the regions are strongly linearly correlated to the industry monthly tractor sales in each of the regions as well.

2ii) How mowers and tractors are performing in terms of gross revenues and market share for each region as well as at the world-wide level. Conduct the appropriate analyses, interpret your results and summarize your findings for her. Incorporate any charts that can help in visualizing the data or results.

To analyse how mowers are performing in terms of gross revenue and market share for each region, contingency tables can be drawn to show the gross revenue and market share based on each year and each region.

To analyse gross revenue of PLE mowers:

For the contingency table below (Table 1), each cell represents the total gross revenue earned with PLE mowers in a particular year and in a particular region. The row totals represent the total gross revenue earned in the particular region in the 5 years, and the column totals represent the total gross revenue earned in the particular year in the world. The cell where the row totals and column totals intersect represents the total gross revenue earned in the world in the 5 years.

Region \ Year	2010	2011	2012	2013	2014	Totals
China	0.00	0.00	0.00	0.00	21,470.00	**21,470.00**
Europe	2,340,000.00	2,688,000.00	2,430,000.00	2,380,950.00	2,205,900.00	**12,044,850.00**
NA	13,575,000.00	15,442,000.00	16,263,000.00	16,946,000.00	17,451,500.00	**79,677,500.00**
Pacific	223,500.00	306,250.00	365,400.00	458,800.00	494,000.00	**1,847,950.00**
SA	439,500.00	549,500.00	610,200.00	667,850.00	735,300.00	**3,002,350.00**
Totals	**16,578,000.00**	**18,985,750.00**	**19,668,600.00**	**20,453,600.00**	**20,908,170.00**	**96,594,120.00**

Table 1: A contingency table of total gross revenues (in dollars) from PLE mowers categorised based on year and region.

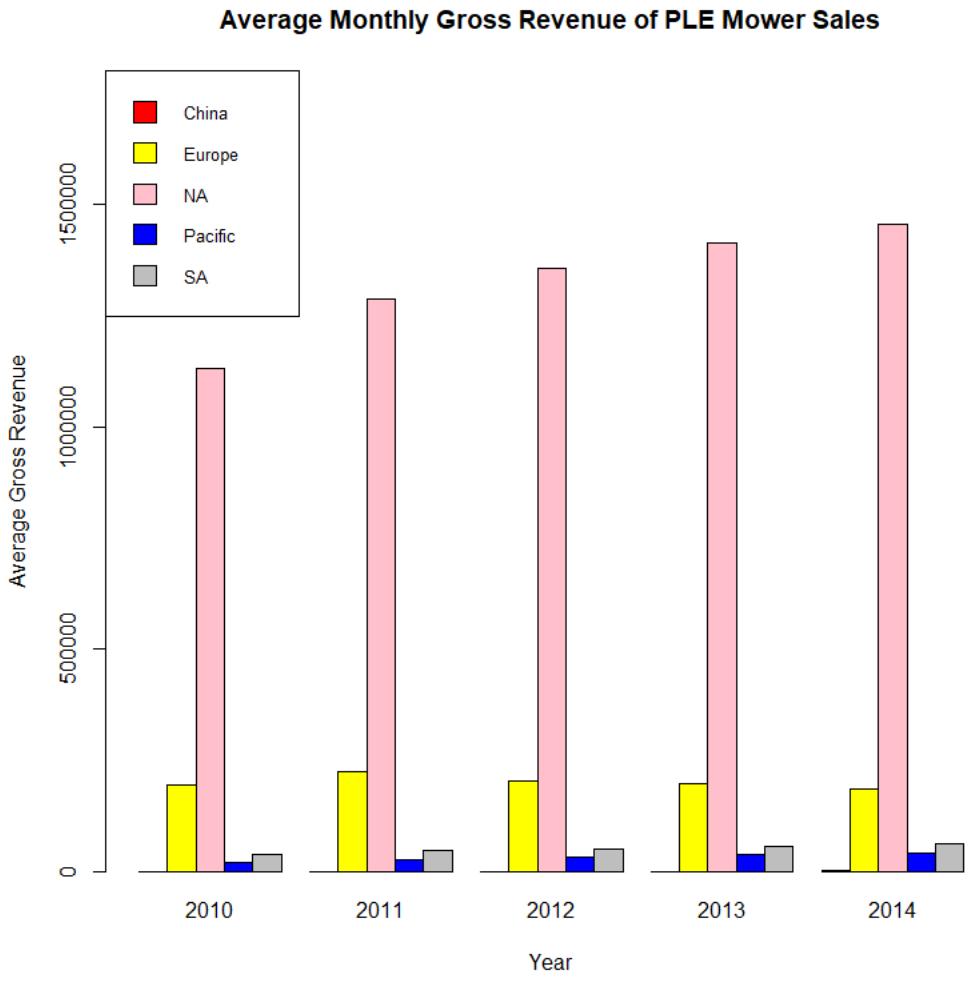
Inference from Table 1:

- World total gross revenue from PLE mowers has been increasing from 2010 to 2014, but the rate of increase has been slowing down over the years. For example, percentage increase from 2010 to 2011 is 14.52% and percentage increase from 2011 to 2012 is 3.60%.
- In the 5 years, PLE has the highest total gross revenue from the mowers market in North America ($79,677,500) and has the lowest total gross revenue from the mowers market in China ($21,470). PLE has only just opened its mower's market in China from 2014, and hence it is understandable that PLE has the lowest total gross revenue from China's market. The next lowest in total gross revenue from the mowers market in the region is the market in Pacific ($1,847,950).
- The highest world total gross revenue from PLE mowers was generated in 2014 ($20,908,170).

For the following contingency table (Table 2), the value of a particular region in a particular year (i.e. in each cell) is the average gross revenue PLE earns per month in that year and in that region. The average gross revenue for each country in a particular year can be seen in the column totals, while the average gross revenue per year in the 5 years for a particular country can be seen in the row totals. The cell where the row totals intersects with the column totals represents the average gross revenue per year for a country.

Region \ Year	2010	2011	2012	2013	2014	Totals
China	0.00	0.00	0.00	0.00	1,789.17	**357.83**
Europe	195,000.00	224,000.00	202,500.00	198,412.50	183,825.00	**200,747.50**
NA	1,131,250.00	1,286,833.33	1,355,250.00	1,412,166.67	1,454,291.67	**1,327,958.33**
Pacific	18,625.00	25,520.83	30,450.00	38,233.33	41,166.67	**30,799.17**
SA	36,625.00	45,791.67	50,850.00	55,654.17	61,275.00	**50,039.17**
Totals	**276,300.00**	**316,429.17**	**327,810.00**	**340,893.33**	**348,469.50**	**321,980.40**

Table 2: A contingency table of average monthly gross revenues (in dollars) from PLE mowers categorised based on year and region.



Graph 29: A clustered bar plot that shows the average monthly gross revenue from PLE mower sales categorized by year and region. Proportion of the average monthly gross revenues of different regions can be clearly seen.

Inference from Graph 29:

- As seen from the grouped bar chart, in each year, the average monthly gross revenue from PLE mower sales is the highest in North America, and the lowest in China.
- The average monthly gross revenues from PLE mower sales in other regions are much lower as compared to that in North America.
- North America's demand is the key source of gross revenue in the market of PLE mowers.

Conclusion:

- PLE mowers market in China is still a developing market, hence the low gross revenue generated.
- Demand for PLE's mowers is the strongest in North America, as compared to the other regions, and hence PLE mowers market in North America is the key source of gross revenue for PLE's mowers.

- PLE is still expanding its mowers markets in the world as evident from the increasing sales each year, but the rate of increase in gross revenue declines as gross revenue from PLE mowers gets higher.

To analyse market share of PLE mowers:

For the table below (Table 3), each cell represents the market share in a particular region, and in a particular year (in %). Market share was calculated by taking the total number of mower units sold in the region by PLE in the year of interest divided by the total number of mower units sold in the region by the industry in the year of interest, multiplied by 100. PLE's market share in the World was calculated by taking the total number of mower units PLE sold in the year divided by the total number of mower units the industry sold in the year, multiplied by 100. The Average column shows the average yearly market share a particular region has.

China is left out of this analysis due to the absence of data about its mower industry total sales.

	2010	2011	2012	2013	2014	Average
North America	10.28	10.34	10.47	10.56	10.32	10.394
South America	35.58	38.91	41.44	45.70	47.40	41.806
Europe	5.55	5.50	5.59	5.68	4.86	5.436
Pacific	10.23	10.51	10.72	10.47	10.89	10.564
World	9.33	9.37	9.66	9.82	9.48	9.532

Table 3: A contingency table that shows the market share that PLE takes up in the mower markets, categorized by year and region.

Inference from Table 3:

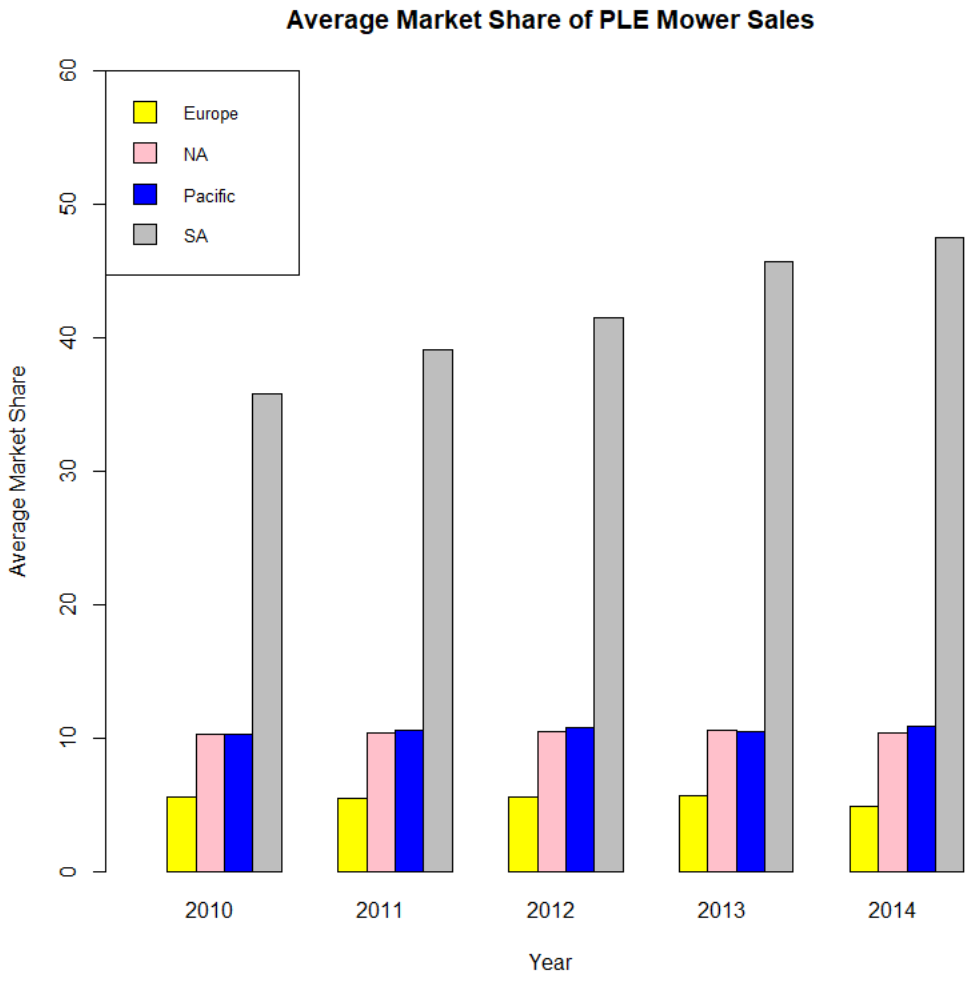
- It is evident that PLE has the highest market share in South America's mowers market (an average yearly market share of 41.81%), as compared to PLE's market shares in other regions.
- PLE has the lowest market share in Europe's mowers market (an average yearly market share of 5.44%), as compared to PLE's market shares in other regions.
- PLE has a similar market share in North America and Pacific's mower markets, with an average yearly market share of 10.39% and 10.56% respectively.
- PLE's market share in the world mowers market was relatively constant in the 5 years, with an average world market share of 9.53%.

For the table below (Table 4), each cell represents the average monthly market share in a particular region, in a particular year. Each cell's value is calculated by taking the sum of the market share of each month of the given year in the given region divided by 12 (number of months) i.e. the market shares of each month of the particular year are calculated and then summed up, and divided by 12). The column totals show the average monthly market share in a region in a particular year; and the row totals show the average monthly market share in a year in a particular region. The cell where the column totals intersect with the row totals represents the average monthly market share in a year in a region (i.e. the average of the row totals).

The average monthly market share is used to estimate the market share of PLE mowers given a region and a year.

Region \ Year	2010	2011	2012	2013	2014	Totals
China						
Europe	5.54	5.47	5.57	5.67	4.87	**5.42**
NA	10.26	10.32	10.48	10.55	10.33	**10.39**
Pacific	10.23	10.53	10.73	10.47	10.89	**10.57**
SA	35.83	39.08	41.50	45.67	47.47	**41.91**
Totals	**15.47**	**16.35**	**17.07**	**18.09**	**18.39**	**17.07**

Table 4: A contingency table that shows the average monthly market share (in %) that PLE takes up in the mower markets, categorized based on year and region.



Graph 30: A clustered bar plot that shows the average monthly market share PLE takes up in the mower markets, categorized by year and region. Proportion of the average monthly market shares of different regions can be clearly seen.

Inference from Graph 30:

- As seen from the grouped bar chart, the average monthly market share PLE takes up in the South American market is the highest, as compared to those of the other regions. The average monthly market share PLE takes up in the European market is the lowest.
- The average monthly market share PLE takes up in the European, North America and Pacific markets remain relatively constant throughout the 5 years.
- The average monthly market share PLE takes up in South America, on the other hand, has been increasing throughout the 5 years.

Conclusion:

- The market share PLE takes up in South America is the highest, although most of PLE's gross revenue comes from the North American market. This may imply that the North American's mowers industry is the largest among all other regions (with the largest demand), and hence most of PLE's mowers were sold to North Americans, resulting in

high gross revenue from North America's sales. However, market share of PLE mowers in the North American mowers industry is low due to the existence of other competition firms in the industry, and thus PLE mowers only take up around 10% of all existing mowers in the North American mowers industry. In contrast, South Americans' industry for mowers might be smaller, but PLE has more market dominance in South America.

- PLE's market share in the South American mowers industry has been rising over the 5 years.
- PLE's market share in the world mowers market was relatively constant in the 5 years, with an average world market share of 9.53%.

We can conduct a similar analysis for PLE tractors.

For the following table below (Table 5), each of the cells represent the total gross revenue (in $) from PLE tractors in a particular year, and in a particular region. This is calculated by taking the total number of PLE tractor units sold in the year, in the region, multiplied by the price of the PLE tractor sold. The column totals represent the total gross revenue gained by PLE tractors in the particular year in the world. The row totals represent the total gross revenue gained by PLE tractors in a particular region, over the 5 years. The cell whereby the row totals and the column totals intersect represents the total gross revenue gained from PLE tractors over the 5 years in the world.

Region \ Year	2010	2011	2012	2013	2014	Totals
China	0.00	0.00	1,047,600.00	3,566,800.00	5,867,200.00	**10,481,600.00**
Europe	26,399,750.00	28,512,400.00	30,240,000.00	27,413,300.00	24,928,000.00	**137,493,450.00**
NA	20,985,250.00	27,557,000.00	36,464,400.00	56,843,100.00	92,902,400.00	**234,752,150.00**
Pacific	9,896,250.00	11,294,800.00	12,358,800.00	14,015,600.00	10,423,400.00	**57,988,850.00**
SA	10,673,000.00	17,533,800.00	26,424,000.00	34,117,700.00	41,416,200.00	**130,164,700.00**
Totals	**67,954,250.00**	**84,898,000.00**	**106,534,800.00**	**135,956,500.00**	**175,537,200.00**	**570,880,750.00**

Table 5: A contingency table of total gross revenues (in dollars) from PLE tractors categorised based on year and region.

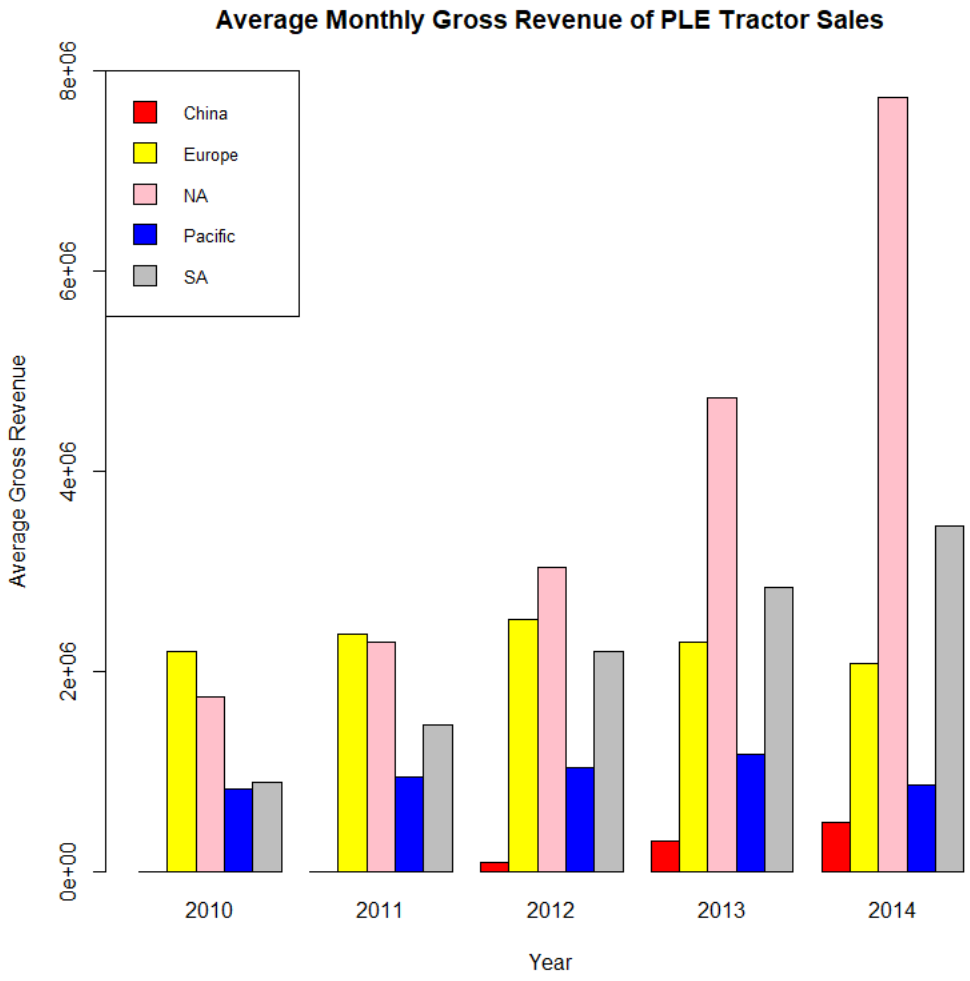
Inference from Table 5:

- In the 5 years, PLE has the highest total gross revenue from the tractors market in North America ($234,752,150) and the lowest total gross revenue from the tractors market in China ($10,481,600) – which was newly opened in 2012.
- World total gross revenue has been increasing over the years, and it has been increasing at an increasing rate. For example, there is a 24.9% increase in world total gross revenue from 2010 to 2011; and there is a 25.5% increase in world total gross revenue from 2011 to 2012 – percentage increase over the years has risen.
- The highest world total gross revenue from PLE tractors was generated in 2014 ($175,537,200), as compared to the world total gross revenue generated in other years.

For the following table below (Table 6), each cell represents the average monthly gross revenue from PLE tractors in a given year and a given region. Each cell's value is calculated by taking the total gross revenue from PLE tractors in the given year and region divided by 12 (number of months). The average gross revenue for each country in a particular year can be seen in the column totals, while the average gross revenue per year in the 5 years for a particular country can be seen in the row totals. The cell where the row totals intersects with the column totals represents the average gross revenue per year for a country.

Region \ Year	2010	2011	2012	2013	2014	Totals
China	0.00	0.00	87,300.00	297,233.33	488,933.33	**174,693.33**
Europe	2,199,979.17	2,376,033.33	2,520,000.00	2,284,441.67	2,077,333.33	**2,291,557.50**
NA	1,748,770.83	2,296,416.67	3,038,700.00	4,736,925.00	7,741,866.67	**3,912,535.83**
Pacific	824,687.50	941,233.33	1,029,900.00	1,167,966.67	868,616.67	**966,480.83**
SA	889,416.67	1,461,150.00	2,202,000.00	2,843,141.67	3,451,350.00	**2,169,411.67**
Totals	**1,132,570.83**	**1,414,966.67**	**1,775,580.00**	**2,265,941.67**	**2,925,620.00**	**1,902,935.83**

Table 6: A contingency table of average monthly gross revenues (in dollars) from PLE tractors categorised based on year and region.



Graph 31: A clustered bar plot that shows the average monthly gross revenue from PLE tractor sales categorized by year and region. Proportion of the average monthly gross revenues of different regions can be clearly seen.

Inference from Graph 31:

- From the grouped bar chart, we can see that in years 2010 and 2011, the average monthly gross revenues from PLE tractor sales were the highest in Europe, but over the 5 years, PLE tractor sales have stayed relatively constant in Europe, and hence Europe was overtaken by PLE tractor sales in North America from 2012 onwards (because the PLE tractor sales in North America have been increasing at an increasing rate over the 5 years).
- China and South American markets are also growing, with increasing average monthly gross revenues from PLE tractor sales over the years; while Europe and Pacific's markets remained relatively the same over the years.
- North America's demand for PLE's tractors forms the key source of gross revenue in the market of PLE tractors.

Conclusion:

- World total gross revenue from the sales of PLE's tractors is growing at an increasing rate, and hence evident that PLE's tractors market in the world is still expanding.
- China and South American markets are also growing over the years, although the rate of growth is not as high as that of the North American markets.
- European and Pacific markets remained relatively the same over the 5 years.
- North America's demand for PLE's tractors forms the key source of gross revenue in the market of PLE tractors.

To analyse market share of PLE tractors:

For the table below (Table 7), each cell is represented by the market share (in %) that PLE tractors take up in the tractors industry in the given year and region. Market share was calculated by taking the total number of tractor units sold in the region by PLE in the year of interest divided by the total number of tractor units sold in the region by the industry in the year of interest, multiplied by 100. PLE's market share in the World was calculated by taking the total number of tractor units PLE sold in the year divided by the total number of tractor units the industry sold in the year, multiplied by 100. The Average column shows the average yearly market share a particular region has.

	2010	2011	2012	2013	2014	Average
North America	7.92	10.74	12.90	15.78	18.70	13.208
South America	26.43	27.34	28.16	29.28	29.68	28.178
Europe	20.87	19.23	16.10	16.67	16.47	17.868
Pacific	21.53	20.88	20.49	20.26	19.80	20.592
World	11.22	12.88	13.94	15.01	16.40	13.890

Table 7: A contingency table that shows the market share that PLE takes up in the tractor markets, categorized by year and region.

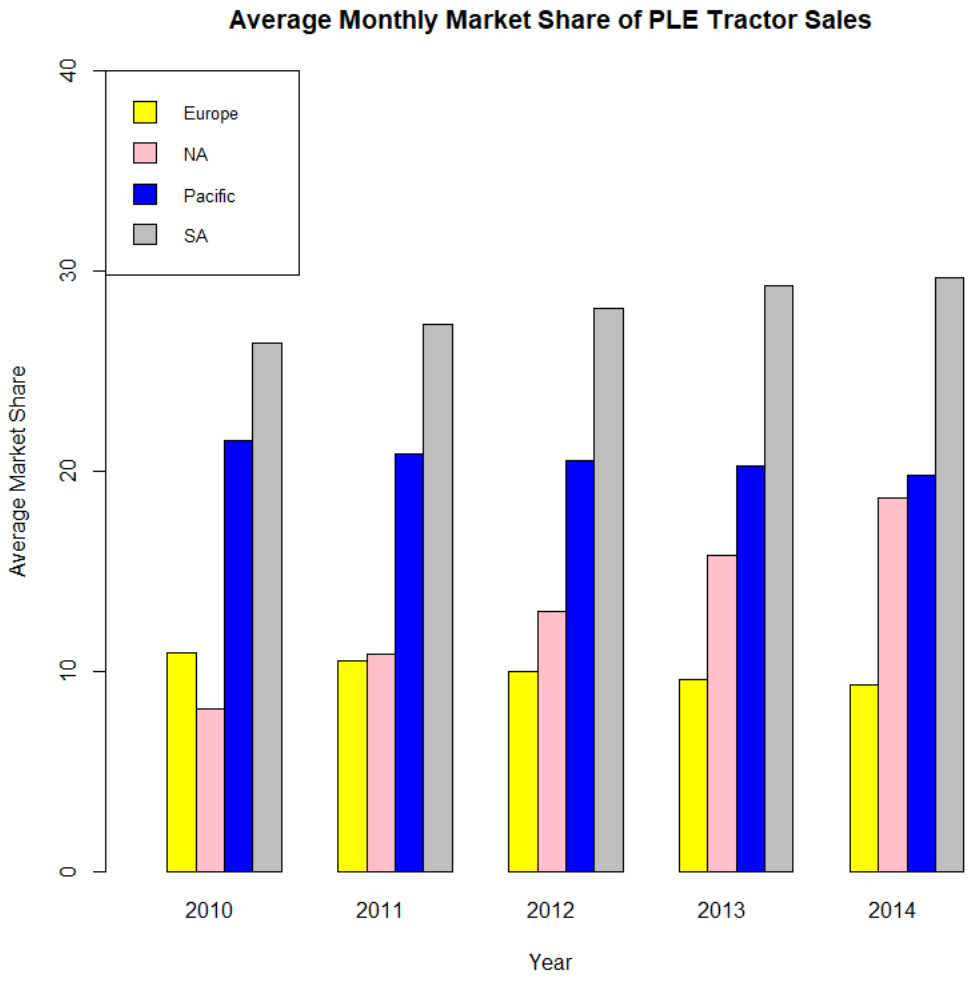
Inference from Table 7:

- It is evident that PLE has the highest market share in South America's tractors market (an average yearly market share of 28.18%), as compared to PLE's market shares in other regions.
- PLE has the lowest market share in North America's tractors market (an average yearly market share of 13.21%), as compared to PLE's market shares in other regions.
- PLE has fairly similar market shares in Europe and Pacific's tractor markets, with an average yearly market share of 17.87% and 20.59% respectively.
- PLE's market share in the world tractors market has been increasing over the 5 years, with an average yearly world market share of 13.89%.

For the table below (Table 8), each cell represents the average monthly market share from PLE tractors in a particular region, given a particular year. Each cell's value is calculated by taking the sum of the market share of each month of the given year in the given region divided by 12 (number of months) i.e. the market shares of each month of the particular year are calculated and then summed up, and divided by 12. The column totals show the average monthly market share in a region in a particular year; and the row totals show the average monthly market share in a year in a particular region. The cell where the column totals intersect with the row totals represents the average monthly market share in a year in a region (i.e. the average of the row totals).

Region \ Year	2010	2011	2012	2013	2014	Totals
China						
Europe	10.90	10.50	10.01	9.60	9.31	**10.06**
NA	8.09	10.83	12.99	15.76	18.65	**13.26**
Pacific	21.52	20.87	20.50	20.26	19.80	**20.59**
SA	26.43	27.35	28.15	29.27	29.68	**28.17**
Totals	**16.73**	**17.39**	**17.91**	**18.72**	**19.36**	**18.02**

Table 8: A contingency table that shows the average monthly market share (in %) that PLE takes up in the tractor markets, categorized based on year and region.



Graph 32: A clustered bar plot that shows the average monthly market share PLE takes up in the tractor markets, categorized by year and region. Proportion of the average monthly market shares of different regions can be clearly seen.

Inference from Graph 32:

- As seen from the grouped bar chart, the average monthly market share PLE takes up in the South American market is the highest, as compared to those of the other regions.
- The average monthly market share PLE tractors take up in the North American and South American markets increased over the 5 years, with North American's rate of increase being higher than that of South America.
- The average monthly market share that PLE tractors take up in the other regions remain fairly constant over the years.

Conclusion:

- Similar to that of mowers, the market share PLE tractors take up in South America is the highest, although most of PLE tractors' gross revenue comes from the North American market.

- PLE's tractor markets in North America and South America have been expanding over the years, with North America increasing at a higher rate.
- PLE's market share in the world tractors market has been increasing over the 5 years, with an average yearly world market share of 13.89%.

3i) Examine and analyze the End-User satisfaction, Dealer satisfaction and Customer Survey data. Provide a detailed statistical summary for the End-User satisfaction, dealer satisfaction and customers' ratings of PLE products. (provide detail breakdown by year and region where applicable).

To analyse the End-User satisfaction ratings, we can use mean (to find the average rating), standard deviation (to find the spread of the data from the mean) and median of the data (to find the middle of the data). Below are tables that summarise these 3 descriptive statistics of the data, categorised by year and region.

Table 9: **North America**'s End-User Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	1	3	6	15	37	38	100	3.98	1.10	4
2	2011	1	2	4	18	35	40	100	4.04	1.04	4
3	2012	1	2	5	17	34	41	100	4.04	1.06	4
4	2013	0	2	4	15	33	46	100	4.17	0.96	4
5	2014	0	2	3	15	31	49	100	4.22	0.95	4

Table 10: **South America**'s End-User Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	1	2	5	18	36	38	100	4.00	1.05	4
2	2011	1	3	6	17	36	37	100	3.95	1.10	4
3	2012	0	2	6	19	37	36	100	3.99	0.99	4
4	2013	0	2	5	20	37	36	100	4.00	0.97	4
5	2014	0	2	5	19	37	37	100	4.02	0.97	4

Table 11: **Europe**'s End-User Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	1	2	4	21	36	36	100	3.97	1.04	4
2	2011	1	2	5	21	34	37	100	3.96	1.06	4
3	2012	1	1	4	26	37	31	100	3.90	0.99	4
4	2013	1	1	3	17	41	37	100	4.07	0.96	4
5	2014	0	1	2	19	45	33	100	4.07	0.83	4

Table 12: **Pacific**'s End-User Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	2	3	5	15	41	34	100	3.92	1.13	4
2	2011	1	2	7	15	41	34	100	3.95	1.06	4
3	2012	1	2	5	16	40	36	100	4.00	1.03	4
4	2013	0	2	4	17	40	37	100	4.06	0.94	4
5	2014	0	1	3	19	42	35	100	4.07	0.87	4

Table 13: **China**'s End-User Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
2	2011	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
3	2012	0	3	3	6	28	10	50	3.78	1.04	4
4	2013	1	2	2	4	30	11	50	3.86	1.07	4
5	2014	0	1	1	3	31	14	50	4.12	0.77	4

Summary of End-User Satisfaction Ratings

Table 14: Table of Summary for **Mean** End-User Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	*NA*	3.97	3.78	3.92	4.00
2	2011	*NA*	3.96	3.92	3.95	3.95
3	2012	3.78	3.90	3.97	4.00	3.99
4	2013	3.86	4.07	4.11	4.06	4.00
5	2014	4.12	4.07	4.11	4.07	4.02

Table 15: Table of Summary for **Standard Deviation** of End-User Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	*NA*	1.04	0.97	1.13	1.05
2	2011	*NA*	1.06	0.85	1.06	1.10
3	2012	1.04	0.99	0.94	1.03	0.99
4	2013	1.07	0.96	1.07	0.94	0.97
5	2014	0.77	0.83	1.09	0.87	0.97

Table 16: Table of Summary for **Median** of End-User Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	NA	4	4	4	4
2	2011	NA	4	4	4	4
3	2012	4	4	4	4	4
4	2013	4	4	4	4	4
5	2014	4	4	4	4	4

To analyse the Dealer Satisfaction ratings, we can use mean (to find the average rating), standard deviation (to find the spread of the data from the mean) and median of the data (to find the middle of the data). Below are tables that summarise these 3 descriptive statistics of the data, categorised by year and region.

Table 17: **North America**'s Dealer Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	1	0	2	14	22	11	50	3.78	0.97	4
2	2011	0	0	2	14	20	14	50	3.92	0.85	4
3	2012	1	1	1	8	34	15	60	3.97	0.94	4
4	2013	1	2	6	12	34	45	100	4.11	1.07	4
5	2014	2	3	5	15	44	56	125	4.11	1.09	4

Table 18: **South America**'s Dealer Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	0	0	0	2	6	2	10	4.00	0.67	4
2	2011	0	0	0	2	6	2	10	4.00	0.67	4
3	2012	0	0	1	4	11	14	30	4.27	0.83	4
4	2013	0	1	1	3	12	33	50	4.50	0.86	5
5	2014	1	1	2	4	22	60	90	4.50	0.91	5

Table 19: **Europe**'s Dealer Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	0	0	1	3	7	4	15	3.93	0.88	4
2	2011	0	0	1	2	8	4	15	4.00	0.85	4
3	2012	0	0	1	2	15	7	25	4.12	0.73	4
4	2013	0	0	1	2	21	6	30	4.07	0.64	4
5	2014	0	0	1	4	17	8	30	4.07	0.74	4

Table 20: **Pacific**'s Dealer Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	0	0	1	2	2	0	5	3.20	0.84	3
2	2011	0	0	1	1	3	0	5	3.40	0.89	4
3	2012	0	0	1	1	3	1	6	3.67	1.03	4
4	2013	0	0	0	2	5	3	10	4.10	0.74	4
5	2014	0	0	1	2	7	2	12	3.83	0.83	4

Table 21: **China**'s Dealer Satisfaction Ratings

	Year	0	1	2	3	4	5	Sample Size	Mean	Standard Deviation	Median
1	2010	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
2	2011	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
3	2012	0	0	0	1	0	0	1	3.00	*NA*	3
4	2013	0	0	1	4	2	0	7	3.14	0.69	3
5	2014	0	0	1	5	8	2	16	3.69	0.79	4

Summary of Dealer Satisfaction Ratings

Table 22: Table of Summary for **Mean** Dealer Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	*NA*	3.93	3.78	3.20	4.00
2	2011	*NA*	4.00	3.92	3.40	4.00
3	2012	3.00	4.12	3.97	3.67	4.27
4	2013	3.14	4.07	4.11	4.10	4.50
5	2014	3.69	4.07	4.11	3.83	4.50

Table 23: Table of Summary for **Standard Deviation** of Dealer Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	*NA*	0.88	0.97	0.84	0.67
2	2011	*NA*	0.85	0.85	0.89	0.67
3	2012	*NA*	0.73	0.94	1.03	0.83
4	2013	0.69	0.64	1.07	0.74	0.86
5	2014	0.79	0.74	1.09	0.83	0.91

Table 24: Table of Summary for **Median** of Dealer Satisfaction Ratings

	Year	China	Europe	North America	Pacific	South America
1	2010	NA	4	4	3	4
2	2011	NA	4	4	4	4
3	2012	3	4	4	4	4
4	2013	3	4	4	4	5
5	2014	4	4	4	4	5

To analyse the 2014 Customer survey, we can use more descriptive statistics, such as with the inclusion of measures such as measures of skewness and kurtosis to look at the shape of the data. We should analyse the survey results by region. Below are tables showing the descriptive statistics by region.

```
Descriptive statistics by group
group: China
            vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
Quality        1 10  3.8 0.92      4    3.88 0.74   2   5     3 -0.43    -0.83 0.29
Ease of Use    2 10  4.1 0.57      4    4.12 0.00   3   5     2  0.07    -0.33 0.18
Price          3 10  3.0 0.67      3    3.00 0.00   2   4     2  0.00    -0.97 0.21
Service        4 10  2.6 0.84      3    2.62 0.74   1   4     3 -0.28    -0.84 0.27
------------------------------------------------------------------------------------------
group: Eur
            vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
Quality        1 30 4.10 0.84      4    4.17 1.48   2   5     3 -0.51    -0.68 0.15
Ease of Use    2 30 4.33 0.66      4    4.42 1.48   3   5     2 -0.44    -0.86 0.12
Price          3 30 3.90 1.09      4    4.08 1.48   1   5     4 -1.19     0.96 0.20
Service        4 30 3.87 1.01      4    4.00 1.48   1   5     4 -0.92     0.48 0.18
------------------------------------------------------------------------------------------
group: NA
            vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
Quality        1 100 4.60 0.65      5    4.70 0.00   1   5     4 -2.21     7.73 0.07
Ease of Use    2 100 4.27 0.83      4    4.41 1.48   1   5     4 -1.59     3.62 0.08
Price          3 100 3.71 1.10      4    3.83 1.48   1   5     4 -0.76    -0.11 0.11
Service        4 100 4.31 0.75      4    4.42 1.48   2   5     3 -0.99     0.82 0.07
------------------------------------------------------------------------------------------


group: Pac
            vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
Quality        1 10  4.4 0.70    4.5    4.50 0.74   3   5     2 -0.56    -1.08 0.22
Ease of Use    2 10  3.9 0.88    4.0    4.00 0.00   2   5     3 -0.73    -0.17 0.28
Price          3 10  4.1 0.57    4.0    4.12 0.00   3   5     2  0.07    -0.33 0.18
Service        4 10  4.3 0.67    4.0    4.38 0.74   3   5     2 -0.31    -1.14 0.21
------------------------------------------------------------------------------------------
group: SA
            vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
Quality        1 50 4.28 0.78      4    4.40 1.48   1   5     4 -1.51     4.02 0.11
Ease of Use    2 50 3.92 0.72      4    3.98 0.00   1   5     4 -1.47     4.29 0.10
Price          3 50 3.50 1.05      4    3.55 1.48   1   5     4 -0.56    -0.48 0.15
Service        4 50 4.24 0.82      4    4.35 1.48   1   5     4 -1.32     2.76 0.12
```

Key: NA – North America, SA – South America, Eur – Europe, Pac – Pacific, China - China

3ii) Analyze if there are any significant differences in customers' ratings of specific product/service attributes (quality, ease of use, price, and service) in the 2014 customer survey?

To compare between customers' ratings of the specific product/service attributes in the 2014 customer survey, we are comparing more than 2 variables (4 variables), and whether the means of each of these variables differ from any of the other means of the other variables.

When comparing multiple variables, we first engage in the Bartlett test to check for the equality of the variances for the different variables.

Bartlett test:
H0: Variances in each of the groups are the same, any difference observed is insignificant.
H1: Variances in each of the groups are different, differences between the variances are significant.

Bartlett test result:

```
        Bartlett test of homogeneity of variances

data:  Value by Group
Bartlett's K-squared = 29.298, df = 3, p-value = 1.939e-06
```

Since p-value from the Bartlett test is 1.939e-06, which is less than 0.05, there is sufficient evidence to reject the null hypothesis that the variances in each of the groups are the same, and to support the alternative hypothesis that differences between the variances of each of these groups are significant.

Although the variances in each of the groups are significantly different, the sample sizes in this case are equal in each of the groups, and hence ANOVA test should still be conducted. (i.e. because if sample sizes are equal, violation of the third assumption – homogeneity of variances between groups – does not have serious impacts.)

We proceed to conduct ANOVA test to compare whether the means of the 4 groups are significantly different from one another.

H0: All means of the 4 groups are the same, any difference observed is insignificant.
H1: At least one mean of the 4 groups is different from the others, differences between the means of the 4 groups are significant.

ANOVA test result:

```
             Df Sum Sq Mean Sq F value   Pr(>F)
Group         3   55.5  18.502   23.69 1.08e-14 ***
Residuals   796  621.6   0.781
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Since p-value is 1.08e-14, which is less than 0.05, we reject the null hypothesis, and conclude that there are significant differences in customers’ ratings over different product/service attributes (namely quality, ease of use, price and service). Mean customer ratings of these product/service attributes are not the same.

4i) Provide detailed descriptive statistics for the employee retention data.

The employee retention data is made up of both numerical variables and categorical variables. Descriptive statistics will have to be done separately for the numerical and categorical variables.

For the numerical variables:
The table below shows the required descriptive statistics of the numerical variables, namely ‘YearsPLE’, ‘YrsEducation’, ‘College GPA’ and ‘Age’.

```
             vars  n  mean   sd median trimmed  mad   min   max range  skew kurtosis   se
YearsPLE        1 40  5.54 2.84   5.60    5.59 2.89  0.30 10.00  9.70 -0.11    -1.04 0.45
YrsEducation    2 40 15.70 1.62  16.00   15.75 1.48 13.00 18.00  5.00 -0.33    -0.84 0.26
College GPA     3 40  2.93 0.46   2.96    2.95 0.30  1.75  3.86  2.11 -0.38     0.43 0.07
Age             4 40 29.12 5.18  28.00   28.84 5.93 22.00 40.00 18.00  0.35    -1.15 0.82
```

Mean and median are descriptive statistics that show the location of the data; standard deviation and range show the dispersion of the data; coefficient of skewness and coefficient of kurtosis show the shape of the data.

Inferences:

- 'Age' has a higher standard deviation than all the other numeric variables, which shows that it has the widest dispersion of data around the mean.
- Coefficient of skewness is negative for left-skewed data and positive for right-skewed data, and thus 'YearsPLE', 'YrsEducation' and 'College GPA' have left-skewed data distribution while 'Age' has a right-skewed data distribution. However, since the absolute values of the coefficient of skewness for each of the variables are <0.5, the data distribution for each of the 4 numeric variables is relatively symmetrical.
- Coefficients of kurtosis are <3 for all 4 numeric variables, which indicates that the data distribution is somewhat flat with a wide degree of dispersion, for all 4 numeric variables.

Alternatively, the summary tables of the variables below will give us the 5 numbers (min, Q1, Q2, Q3, max) and the mean of each variable.

```
    YearsPLE       YrsEducation    College GPA         Age
 Min.   : 0.300   Min.   :13.00   Min.   :1.750   Min.   :22.00
 1st Qu.: 3.700   1st Qu.:15.00   1st Qu.:2.750   1st Qu.:25.00
 Median : 5.600   Median :16.00   Median :2.955   Median :28.00
 Mean   : 5.537   Mean   :15.70   Mean   :2.934   Mean   :29.12
 3rd Qu.: 7.675   3rd Qu.:16.25   3rd Qu.:3.150   3rd Qu.:34.00
 Max.   :10.000   Max.   :18.00   Max.   :3.860   Max.   :40.00
```

The third quartile minus the first quartile of the data would give us the interquartile range of each variable, which is also a descriptive statistic that shows us the dispersion of the data.

Interquartile range of each of the variables:
'YearsPLE': 3.975
'YrsEducation': 1.25
'College GPA': 0.4
'Age': 9

For the categorical variables:

We will represent the descriptive statistics of categorical variables through frequency tables. The count of the categorical variables can be seen from the frequency table; modes for each variable can also be found by finding the category of the highest frequency. Mode is a descriptive statistic that shows the location of the data.

Gender

```
  Gender freq
1      F   13
2      M   27
```

College Graduation Status

```
  College.Grad freq
1            N   13
2            Y   27
```

Locality

```
  Local freq
1     N   17
2     Y   23
```

Mode for each of the categorical variables above is 'Y'.

Is there any significant differences in employee retention due to gender, college graduation status, or whether the employee is from the local area?

To check whether there are significant differences in employee retention due to gender, college graduation status and whether the employee is from the local area, we have to run a series of hypothesis tests, one for each variable listed.

Gender

H0: Difference in mean employee retention of females and mean employee retention of males is insignificant (equal to 0).
H1: Difference in mean employee retention of females and mean employee retention of males is significant (not equal to 0).

t-test result:

```
        Welch Two Sample t-test

data:  subset1$YearsPLE and subset2$YearsPLE
t = -0.0091748, df = 18.311, p-value = 0.9928
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.290560  2.270617
sample estimates:
mean of x mean of y
 5.530769  5.540741
```

Therefore, since the p-value from the t-test is 0.9928, which is greater than 0.05, there is insufficient evidence to reject the null hypothesis that the difference between the mean employee retention of males and the mean employee retention of females is not significant. Hence, the null hypothesis stands and mean employee retention of males and mean employee retention of females are the same (difference is insignificant).

College Graduation Status

H0: Difference in mean employee retention of college graduates and mean employee retention of non-college graduates is insignificant (equal to 0).

H1: Difference in mean employee retention of college graduates and mean employee retention of non-college graduates is significant (not equal to 0).

t-test result:

```
        Welch Two Sample t-test

data:  subset1$YearsPLE and subset2$YearsPLE
t = 1.0788, df = 29.24, p-value = 0.2895
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8556063  2.7672872
sample estimates:
mean of x mean of y
 5.848148  4.892308
```

Therefore, since the p-value from the t-test is 0.2895, which is greater than 0.05, there is insufficient evidence to reject the null hypothesis that the difference between the mean employee retention of college graduates and the mean employee retention of non-college graduates is not significant. Hence, the null hypothesis stands and mean employee retention of college graduates and mean employee retention of non-college graduates are the same (difference is insignificant).

Locality

H0: Difference in mean employee retention of locals and mean employee retention of non-locals is insignificant (equal to 0).
H1: Difference in mean employee retention of locals and mean employee retention of non-locals is significant (not equal to 0).

t-test result:

```
        Welch Two Sample t-test

data:  subset1$YearsPLE and subset2$YearsPLE
t = 4.4186, df = 33.866, p-value = 9.674e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.791970 4.844858
sample estimates:
mean of x mean of y
 6.947826  3.629412
```

Since p-value 9.674e-05 is less than 0.05, there is sufficient evidence to reject the null hypothesis that the difference between the mean employee retention of locals and the mean employee retention of non-locals is not significant. Hence, the null hypothesis is rejected and we can conclude that the mean employee retention of locals and mean employee retention of non-locals are not the same (difference is significant).

Conclusion:
There are significant differences in employee retention between employees from the local area and employees not from the local area. The other factors such as gender and whether the

employee is a college graduate are not associated with significant differences in employee retention.

5i) For the *Mower Test* data:
a. What fraction of mowers fails the functional performance test (using all the mower test data)?

54 out of 3000 mowers failed the functional performance test, and hence the fraction of mowers that failed the test is 0.018 or 9/500.

b. What would be a prediction interval for an additional sample of mower test performance?

To look at the prediction interval for an additional sample of mower test performance, numeric data is required, hence, we will have to change the categorical results of the performance ("Pass", "Fail") to numeric representations. In doing so, we assign the numbers "1" and "0" to "Pass" and "Fail" respectively.

We will then compute the 95% prediction interval for the mean of an additional sample of mower test performance.

We will use the following formula to compute the 95% prediction interval:

$$\bar{x} \pm t_{\alpha/2,n-1}\left(s\sqrt{1+\frac{1}{n}}\right)$$

Sample mean = 0.982
Sample standard deviation = 0.01242
Sample size = 30
Alpha = 0.05
t-value = 2.0452
Lower-bound prediction interval = 0.956
Upper-bound prediction interval = 1.008

The prediction interval for the mean of an additional sample of mower test performance is [0.956, 1.008].

This range predicts the value of the mean of an additional sample of mower test performance. There is a 95% probability that the value of the new observation will be in this range.

5ii) For the data in *Blade Weight* worksheet:
a. What is the average blade weight and how much variability is occurring in the measurements of blade weights?

Using the describe() function in R, we obtain the following descriptive statistics of the measurements of blade weights.

```
   vars   n mean   sd median trimmed mad  min  max range skew kurtosis   se
X1    1 350 4.99 0.11   4.99    4.99 0.1 4.63 5.87  1.24 1.41    11.12 0.01
```

Average blade weight is 4.99 units.

To answer how much variability is occurring in the measurements of blade weights, we look at the measures of dispersion to describe the variability of the data.

The measures of dispersion we will use are the range, standard deviation, interquartile range and the coefficient of variation of the dataset.

Range of blade weights = 5.87 – 4.63 = 1.24 units
Standard deviation of blade weights = 0.11 units
Interquartile range of blade weights = 0.14 units
Coefficient of Variation of blade weights = 0.11/4.99 = 0.0220 units (3s.f.)

Interpretation:

- Range of blade weights is small, indicating less dispersion in the data.
- Standard deviation of the blade weights is small, and hence this implies a narrow dispersion of blade weight measurements around the mean. Using the second empirical rule, we can expect 95% of the blade weight measurements to be within the range of [4.88, 5.1], which is a fairly narrow range.
- Interquartile range of blade weights is small, indicating a narrow spread of data.
- Coefficient of Variation is low as well (with CV < 1), and hence we can see that the dataset has low variability.

b. What is the probability of blade weights exceeding 5.2, assuming that the data is normally distributed? Is normal distribution the right assumption for this data?

Assuming that the data for blade weights is normally distributed, the probability of blade weights exceeding 5.2 units is 0.0278 (3s.f.).

To check if normal distribution is the right assumption for this data, we used the Shapiro-Wilkins test.

Test result:

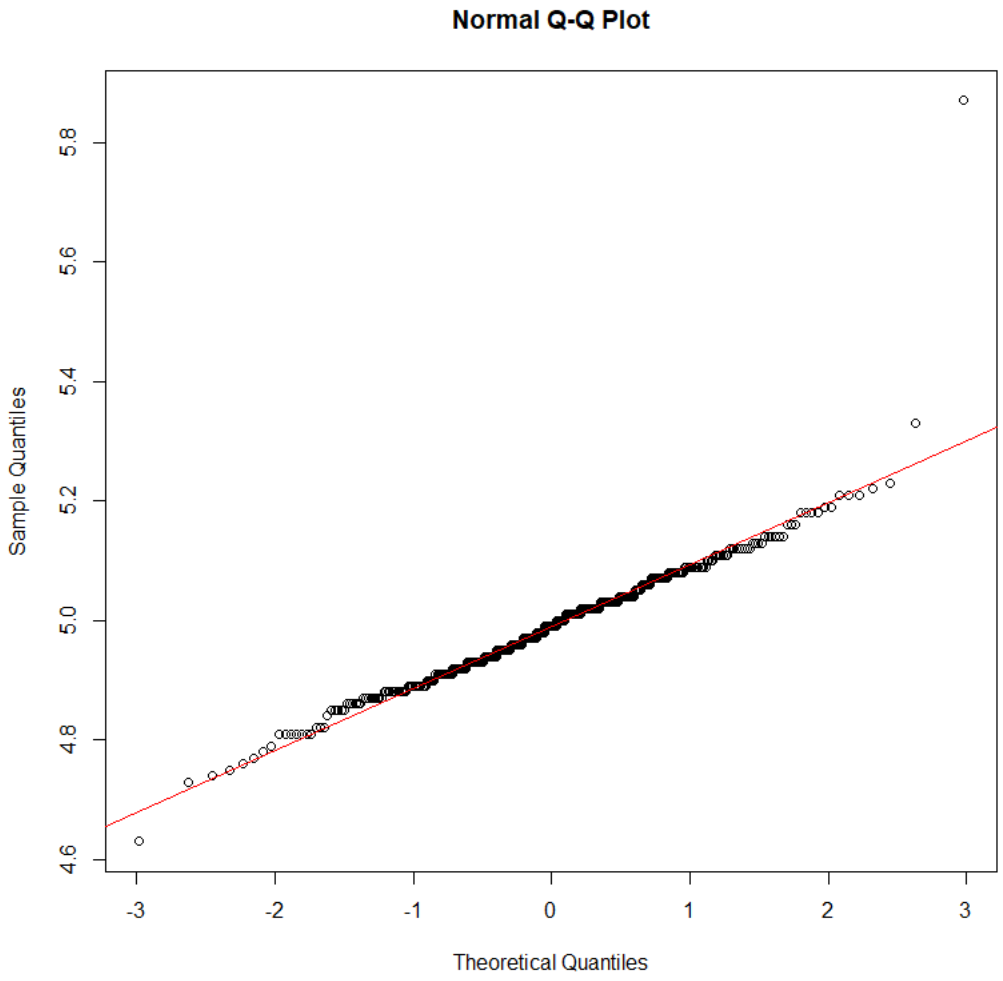
```
        Shapiro-Wilk normality test

data:  blade_weight$Weight
W = 0.92013, p-value = 1.063e-12
```

Since p-value from the Shapiro-Wilk normality test is 1.063e-12, which is less than 0.05, the distribution of the data is significantly different from the normal distribution at 5% level of significance, and hence we cannot assume that the data follows a normal distribution.

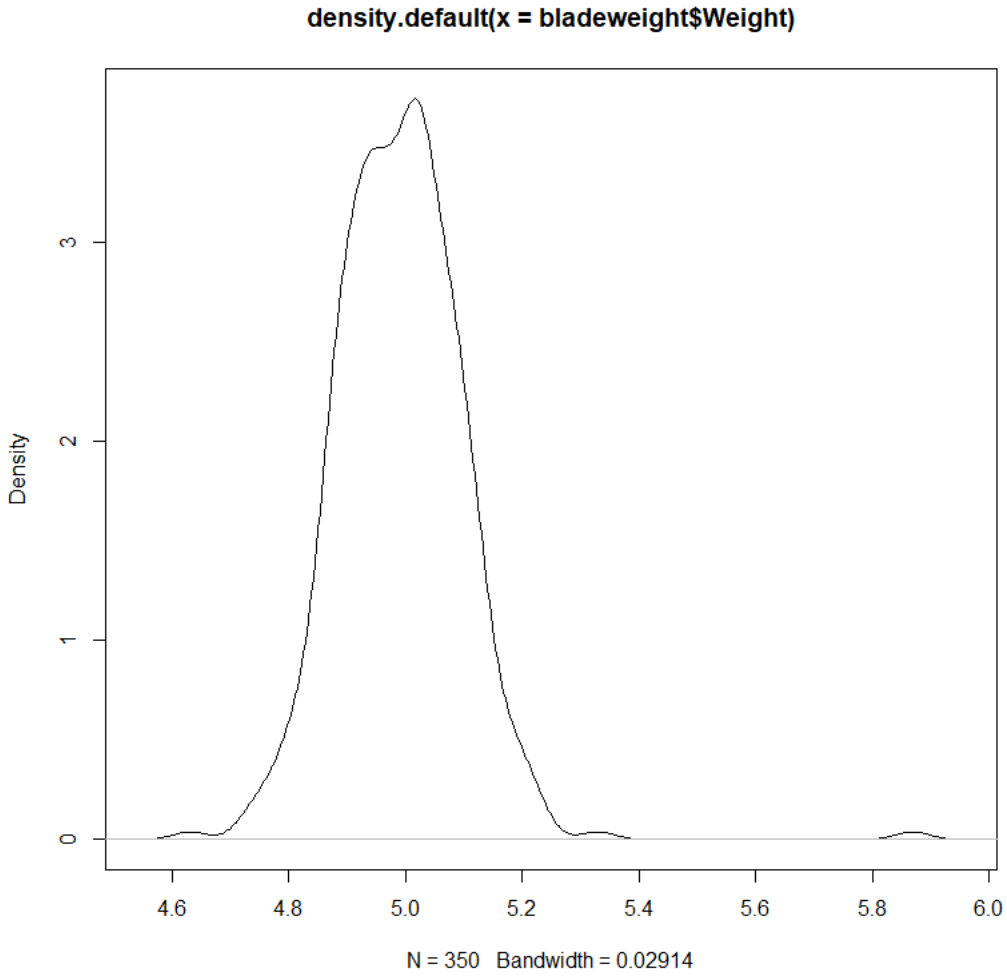
To further check if the dataset follows a normal distribution, a Q-Q plot, a density plot and a histogram can be plotted to visualise the normality of the dataset.

Q-Q Plot



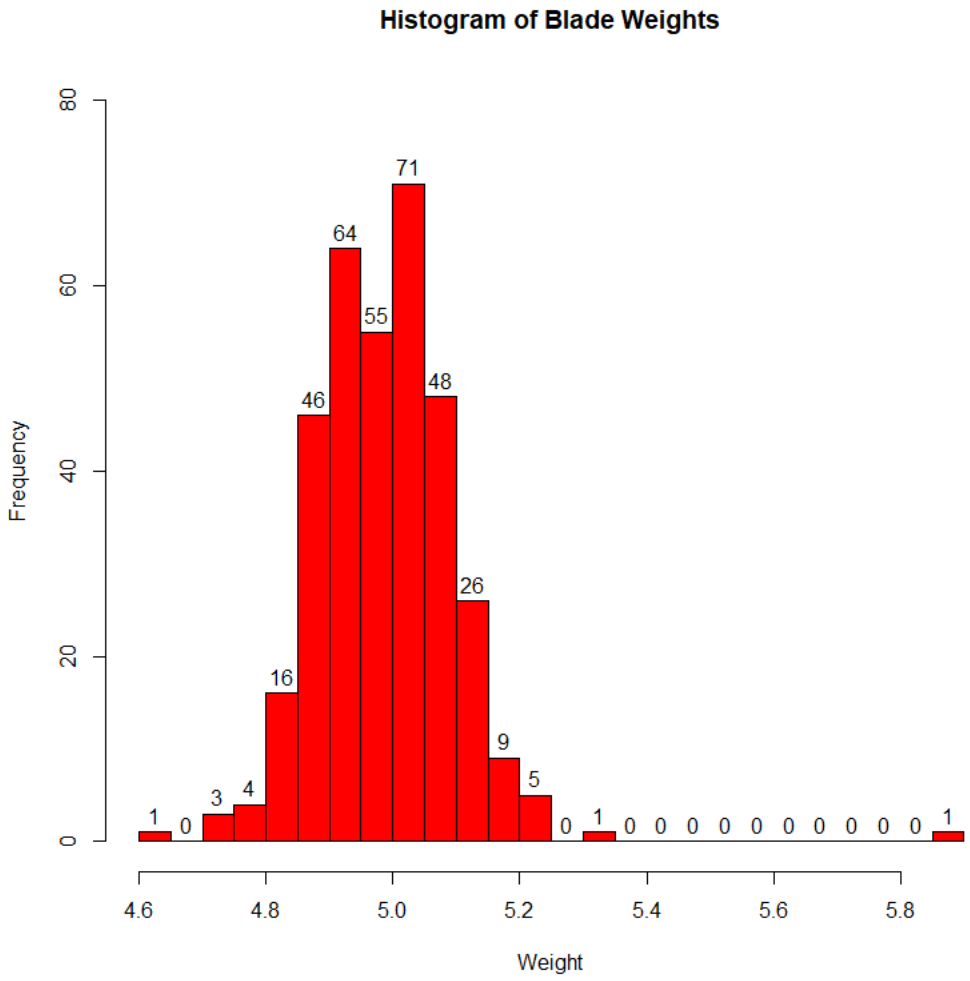
Plot 1: Q-Q plot of blade weights.

Density Plot



Plot 2: Density plot of blade weights.

Histogram Plot



Plot 3: Histogram of blade weights.

Based on the plots above, we can see that the blade weight data that we have does not follow a normal distribution as the data is not symmetrical about the mean. We can also see that the dataset has a right-skewed distribution. Hence, we can conclude that normal distribution is not the right assumption to make for this dataset.

c. What is the sampling distribution of the mean, the overall mean, and the standard error of the mean? Is a normal distribution an appropriate assumption for the sampling distribution of the mean?

The central limit theorem states that if the sample size is large enough, the sampling distribution of the mean is approximately normally distributed, regardless of the distribution of the population and that the mean of the sampling distribution will be the same as that of the population.

By the Central Limit Theorem, the sampling distribution of the mean is approximately normal, since the sample size = 350, which is large enough (>50).

Let X be the weight of the blade.
$\bar{X}$ ~ N(4.991, 3.413e-05) is the sampling distribution of the mean of X.

The overall mean blade weight is 4.99 units.

To calculate the standard error of the mean of the data, we use the following formula:

$$\text{Standard Error of the Mean} = \sigma / \sqrt{n}$$

Using the data provided, and assuming sample standard deviation of blade weights is approximately equals to the population standard deviation of blade weights:

Sample standard deviation of blade weights = 0.11 units
Sample size = 350
Standard error of the mean = 0.00584 (3s.f.)

By Central Limit Theorem, normal distribution is an appropriate assumption for the sampling distribution of the mean, since the sample size is large enough.

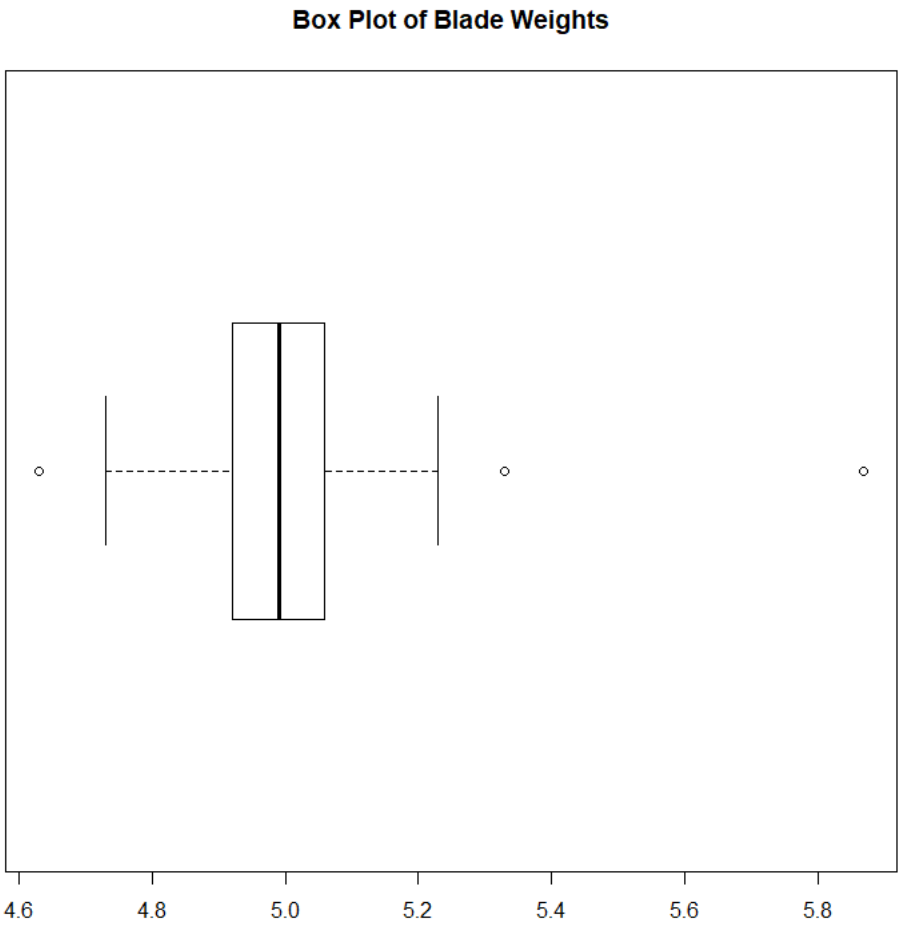
d. Examine the pattern of the blade weight data. Is there any evidence or suggestion of a possible problem with the manufacturing process?

To check if there are any possible problems with the manufacturing process, we will conduct an outlier analysis of the blade weight data to detect any anomalies in the manufacturing process.

To conveniently obtain the five numbers of the data (minimum, lower quartile (Q1), median (Q2), upper quartile (Q3), maximum), we can use the fivenum() function in R.

The five numbers of the blade weight data are 4.63, 4.92, 4.99, 5.06, 5.87. We can use these numbers for our outlier analysis afterwards.

Firstly, we can plot a box-plot to examine if there are any outliers in the blade weight data.



Plot 4: Box-plot of blade weights.

From the box-plot, we can see that there are 3 outliers, represented by the circles in the diagram. There are 2 on the right, which exceeded the upper fence of our data (1.5*IQR + Q3) and 1 on the left is lower than the lower fence of our data (Q1 – 1.5*IQR).

Upon knowing that there are outliers, we can proceed to identify mild and extreme outliers.

Using R, we have identified this sample unit to be an extreme outlier beyond the upper fence.

```
  Sample Weight
   <dbl>  <dbl>
1    171   5.87
```

We have also identified the 2 other mild outliers.

This sample unit is a mild outlier beyond the upper fence.

```
  Sample Weight
   <dbl>  <dbl>
1    172   5.33
```

This sample unit is a mild outlier beyond the lower fence.

```
  Sample Weight
   <dbl>  <dbl>
1     37   4.63
```

With 3 outliers present in the data set (out of 350 other sampling units), the percentage of outliers' occurrence is low (0.857%), and thus there might not be enough evidence to conclude a possible problem with the manufacturing process, especially seeing that there is only 1 extreme outlier. But, this depends on the level of tolerance of outliers in the given manufacturing industry.

e. How many blade weights must be measured to find a 95% confidence interval for the mean blade weight with a sampling error of at most 0.2? What if the sampling error is specified as 0.1?

To calculate the number of blade weights that must be measured to find a 95% confidence interval for the mean blade weight with a sampling error of at most 0.2, we will use the following formula:

$$n \geq (z_{\alpha/2})^2 \frac{(\sigma^2)}{E^2}$$

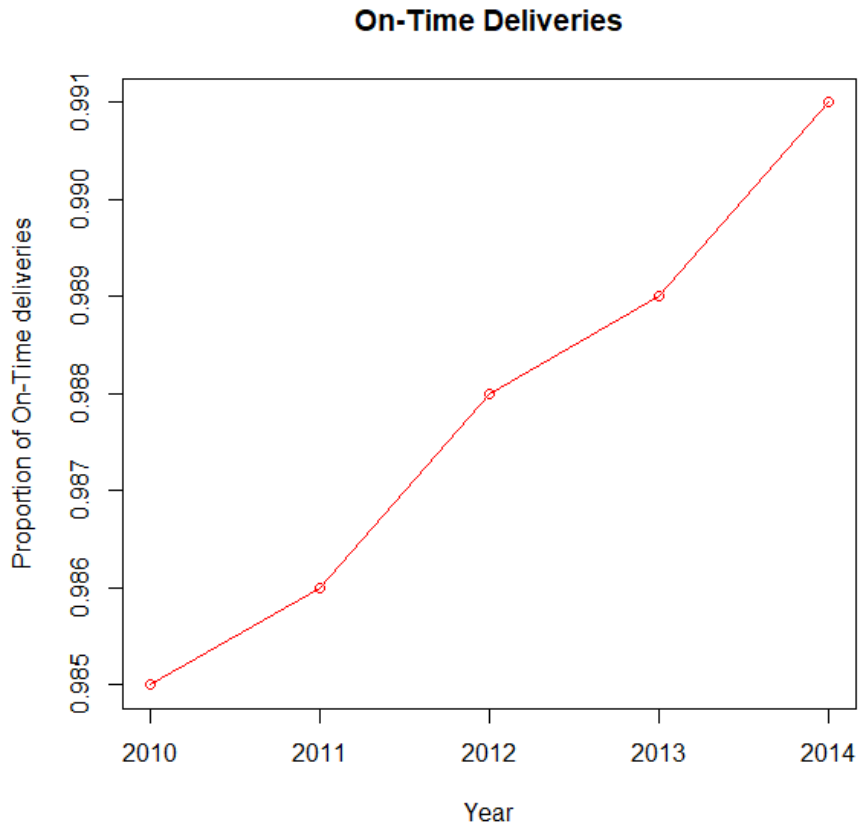
At 95% confidence interval, alpha = 0.05, and hence the z-value = 1.96.
E = 0.2
Thus, $n \geq (1.96)^2 ((0.10928^2)/(0.2^2))$
Simplified, $n \geq 1.14$, and hence at least 2 blade weights must be measured to find a 95% confidence interval for the mean blade weight with a sampling error of at most 0.2.

If the sampling error is specified as 0.1, since the sampling error is reduced, we will need a larger number of samples.

At 95% confidence interval, alpha = 0.05, and hence the z-value = 1.96.
E = 0.1
Thus, $n \geq (1.96)^2 ((0.10928^2)/(0.1^2))$
Simplified, $n \geq 4.58$, and hence at least 5 blade weights must be measured to find a 95% confidence interval for the mean blade weight with a sampling error of at most 0.1.

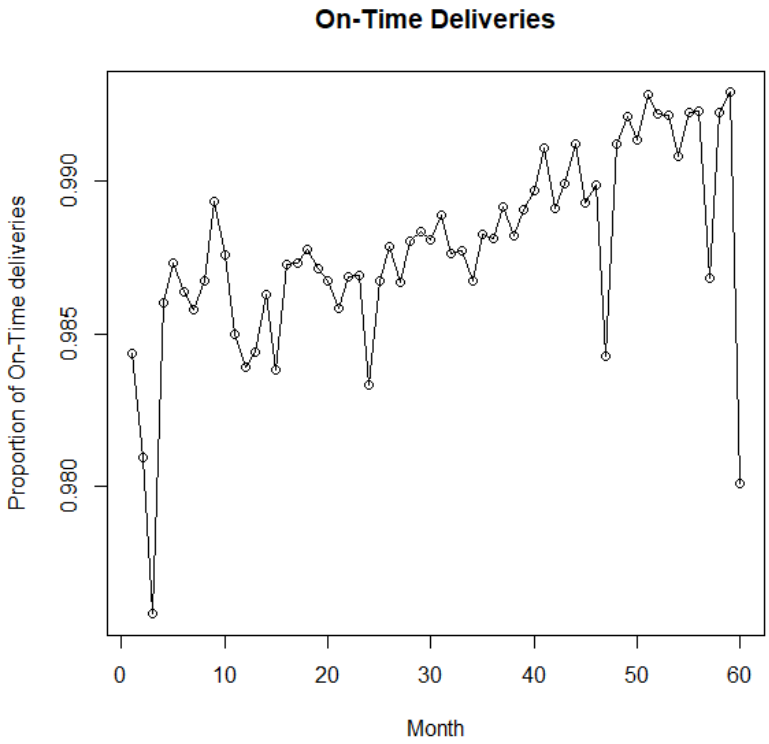
6i) Chart the proportion of on-time deliveries from 2010 to 2014 using the monthly data in the *On-time Delivery* worksheet. Has the proportion of on-time deliveries in the last two years (i.e.2013 & 2014) significantly improved compared to 2010?

To chart the proportion of on-time deliveries from 2010 to 2014, we can plot a line graph with the mean proportion of on-time deliveries in each year, from year 2010 to year 2014.

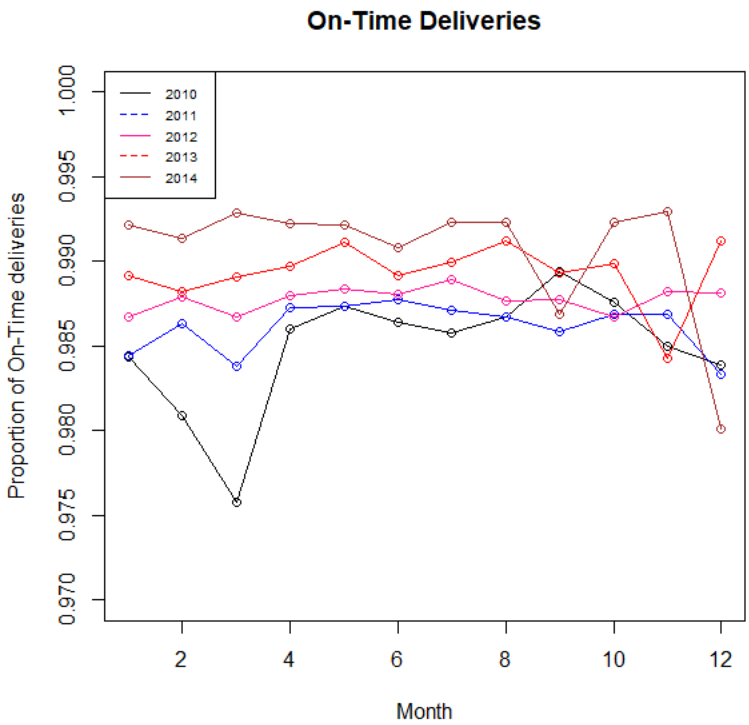


Graph 33: A line graph of the proportion of on-time deliveries, plotted against the year the data was collected in.

Based on monthly data, we can also plot the following line graphs directly.

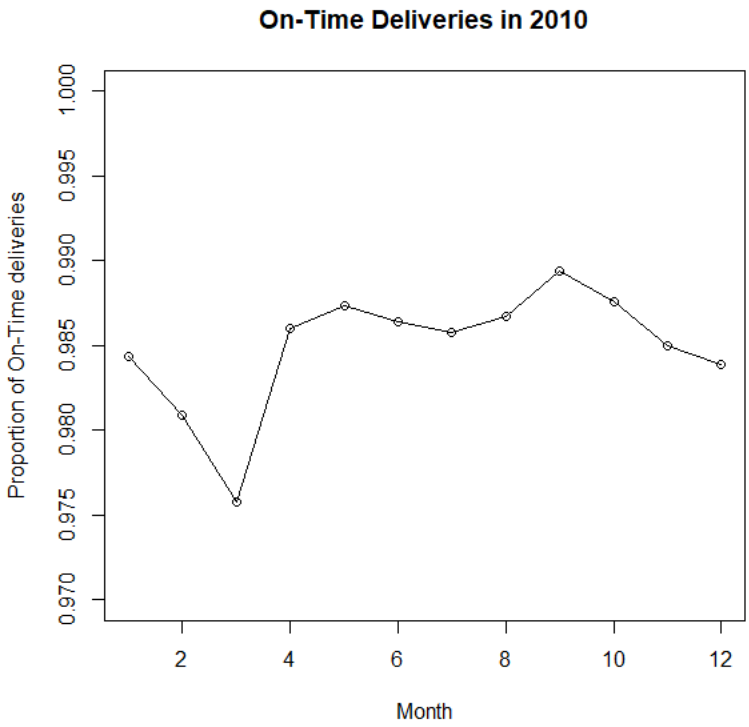


Graph 34: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in, with Month 1 being Jan 2010 and Month 60 being Dec 2014.

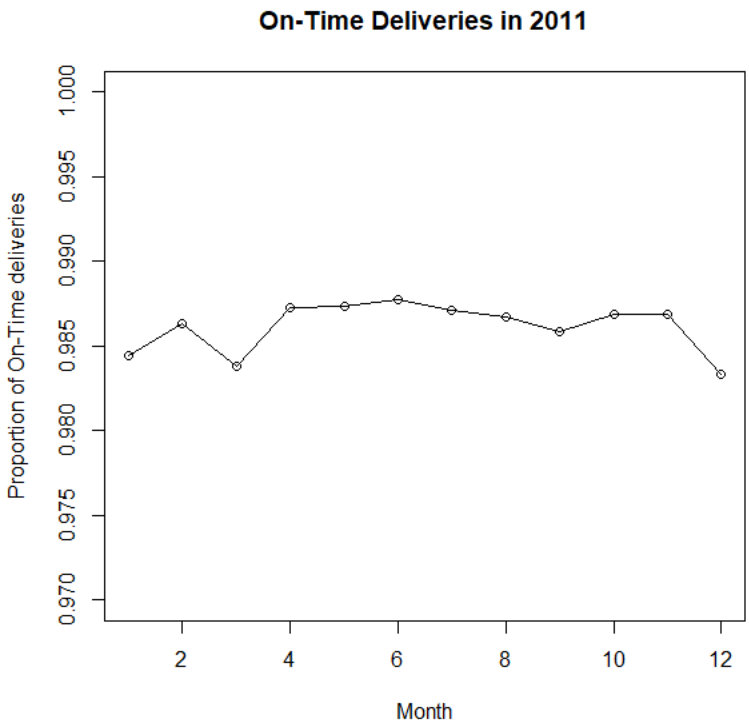


Graph 35: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Each data series (each line) represents a year in the 5 years from 2010 to 2014 from which the data was collected in.

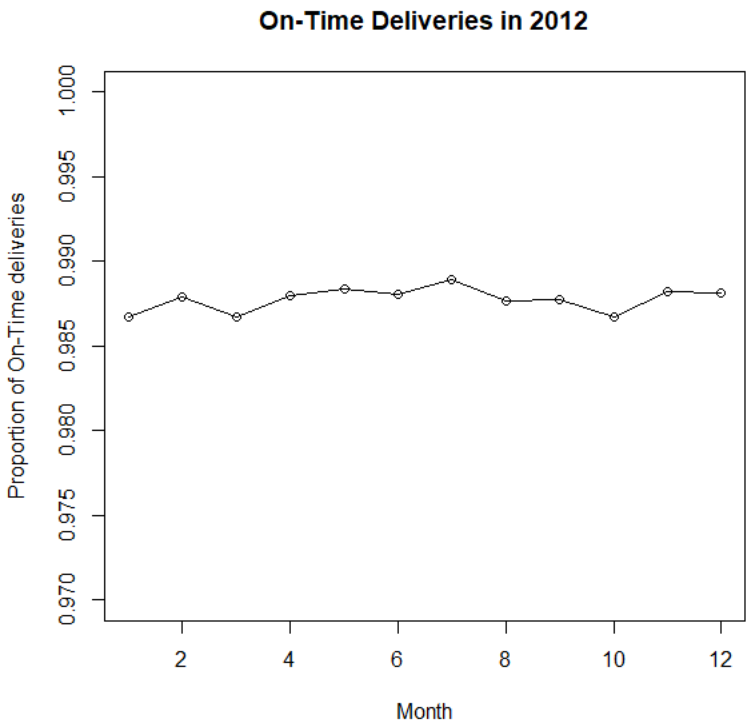
A line chart for each data series:



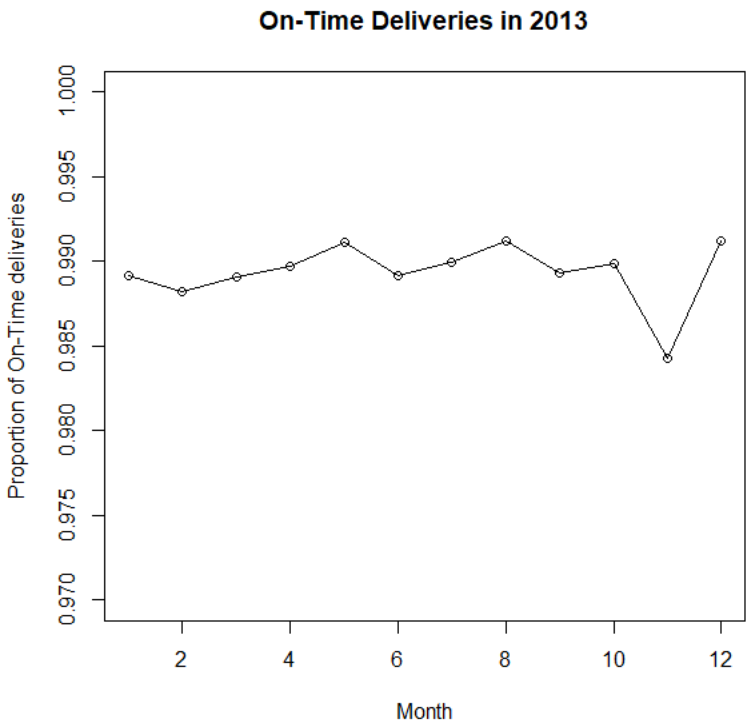
Graph 35: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Data used was collected in the year 2010.



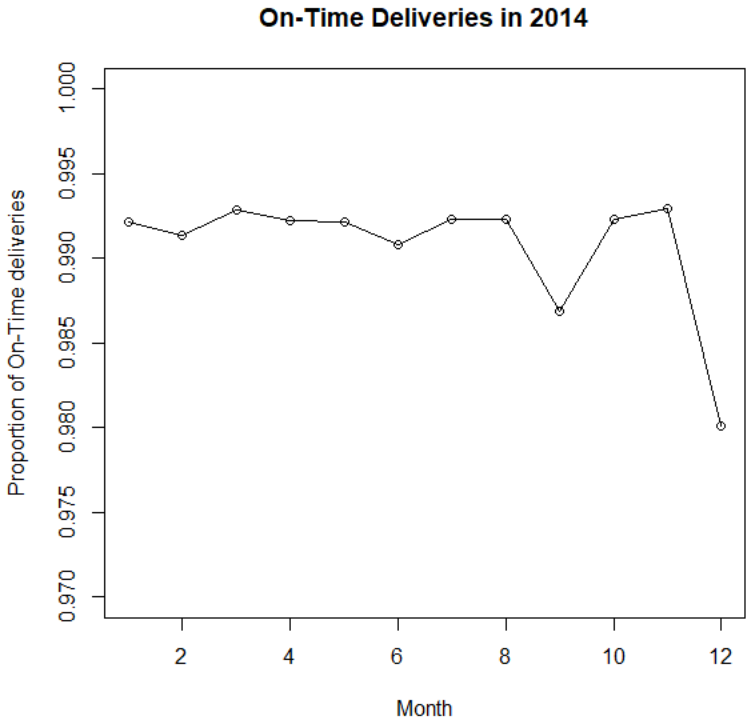
Graph 36: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Data used was collected in the year 2011.



Graph 37: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Data used was collected in the year 2012.



Graph 38: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Data used was collected in the year 2013.



Graph 39: A line graph of the proportion of on-time deliveries, plotted against the month the data was collected in. Data used was collected in the year 2014.

To check whether the proportion of on-time deliveries in year 2013 and year 2014 has significantly improved compared to year 2010, we can conduct two-sample hypothesis tests.

Firstly, we conduct a two-sample hypothesis test on whether the mean proportion of on-time deliveries has significantly improved in year 2013, compared to year 2010.

H0: Mean proportion of on-time deliveries in 2013 – Mean proportion of on-time deliveries in 2010 ≤ 0
H1: Mean proportion of on-time deliveries in 2013 – Mean proportion of on-time deliveries in 2010 > 0

Test result:

```
        Welch Two Sample t-test

data:  ontimedelivery$Percent[c(1:12, 37:48)] by ontimedelivery$Year[c(1:12,
37:48)]
t = -3.804, df = 16.574, p-value = 0.9993
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.006453471          Inf
sample estimates:
mean in group 2010 mean in group 2013
         0.9849328          0.9893591
```

Conclusion:
At 5% level of significance, alpha = 0.05, and thus since p-value = 0.9993, which is greater than 0.05, we cannot reject the null-hypothesis that the mean proportion of on-time deliveries in 2013 did not significantly improve from the mean proportion of on-time deliveries in 2010. Null hypothesis that the true difference in the means is less than or equal to 0 still stands.

Secondly, we conduct a two-sample hypothesis test on whether the mean proportion of on-time deliveries has significantly improved in year 2014, compared to year 2010.

H0: Mean proportion of on-time deliveries in 2014 – Mean proportion of on-time deliveries in 2010 ≤ 0
H1: Mean proportion of on-time deliveries in 2014 – Mean proportion of on-time deliveries in 2010 > 0

Test result:

```
        Welch Two Sample t-test

data:  ontimedelivery$Percent[c(1:12, 49:60)] by ontimedelivery$Year[c(1:12,
49:60)]
t = -3.8688, df = 21.968, p-value = 0.9996
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.008309405          Inf
sample estimates:
mean in group 2010 mean in group 2014
         0.9849328          0.9906877
```

Conclusion:
At 5% level of significance, alpha = 0.05, and thus since p-value = 0.9996, which is greater than 0.05, we cannot reject the null-hypothesis that the mean proportion of on-time deliveries in 2014 did not significantly improve from the mean proportion of on-time deliveries in 2010. Null hypothesis that the true difference in the means is less than or equal to 0 still stands.

Thirdly, we can conduct a two-sample hypothesis test on whether the mean proportion of on-time deliveries has significantly improved in year 2014, compared to year 2013.

H0: Mean proportion of on-time deliveries in 2014 – Mean proportion of on-time deliveries in 2013 ≤ 0
H1: Mean proportion of on-time deliveries in 2014 – Mean proportion of on-time deliveries in 2013 > 0

Test result:

```
        Welch Two Sample t-test

data:  ontimedelivery$Percent[c(37:48, 49:60)] by ontimedelivery$Year[c(37:4
8, 49:60)]
t = -1.108, df = 16.216, p-value = 0.858
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.003420583          Inf
sample estimates:
mean in group 2013 mean in group 2014
         0.9893591          0.9906877
```

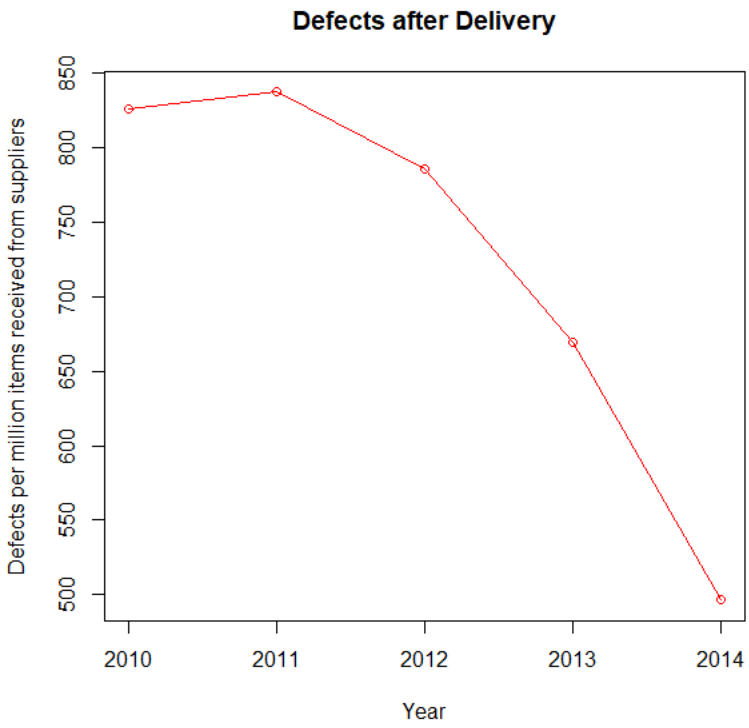
Conclusion:

At 5% level of significance, alpha = 0.05, and thus since p-value = 0.858, which is greater than 0.05, we cannot reject the null-hypothesis that the mean proportion of on-time deliveries in 2014 did not significantly improve from the mean proportion of on-time deliveries in 2013. Null hypothesis that the true difference in the means is less than or equal to 0 still stands.

Hence, we can see that the mean proportion of on-time deliveries did not improve significantly in 2013 from 2010, and in 2014 from 2010. One reason why both results were similar (that there are no significant improvements) for both years 2013 and 2014 is because the mean proportions of on-time deliveries for years 2013 and 2014 are not significantly different from each other as well.

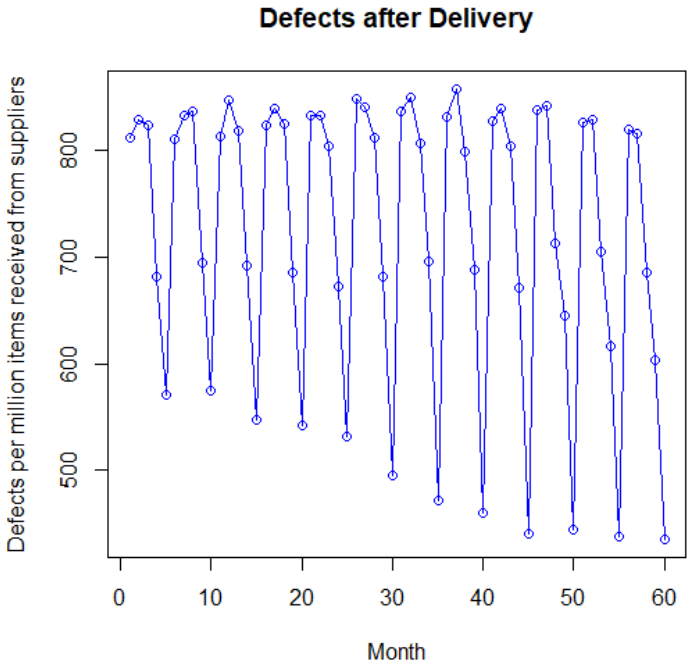
6ii) Chart the number of Defects After Delivery from 2010 to 2014 using the monthly data in the *Defects After Delivery* worksheet.

To chart the number of Defects After Delivery, we can use the monthly data to generate the mean number of defects after delivery in each year from 2010 to 2014. We can then plot a line graph of the mean number of defects after delivery each year against the year the data was collected in.

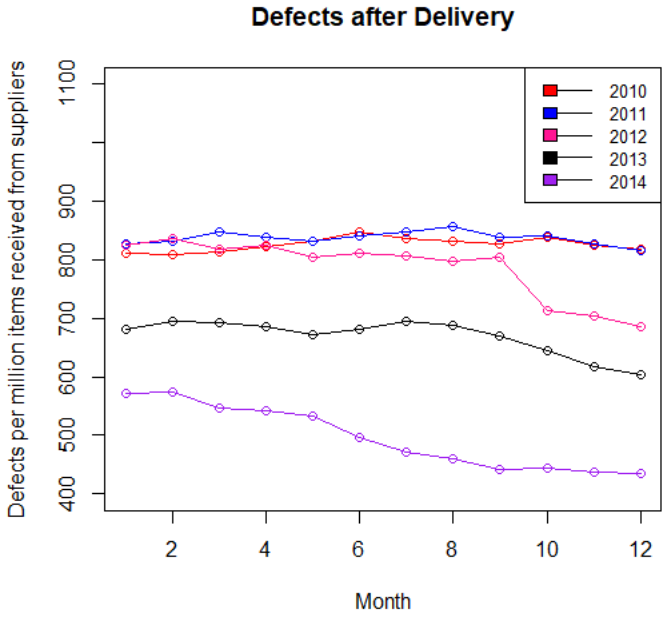


Graph 40: Line graph of the mean number of defects after delivery, plotted against the year the data was collected in.

Based on the monthly data given, we can also plot the following line graphs directly.

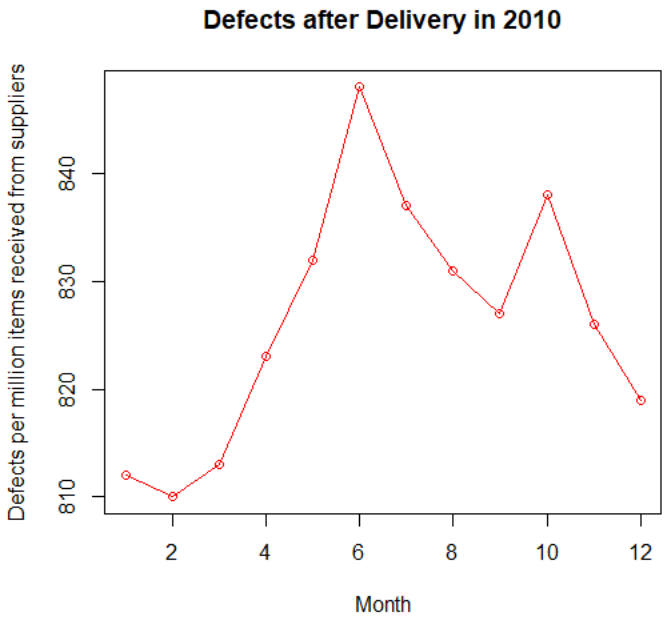


Graph 41: Line graph of the number of defects after delivery, plotted against the month the data was collected in, month 1 being Jan 2010 and month 60 being Dec 2014.

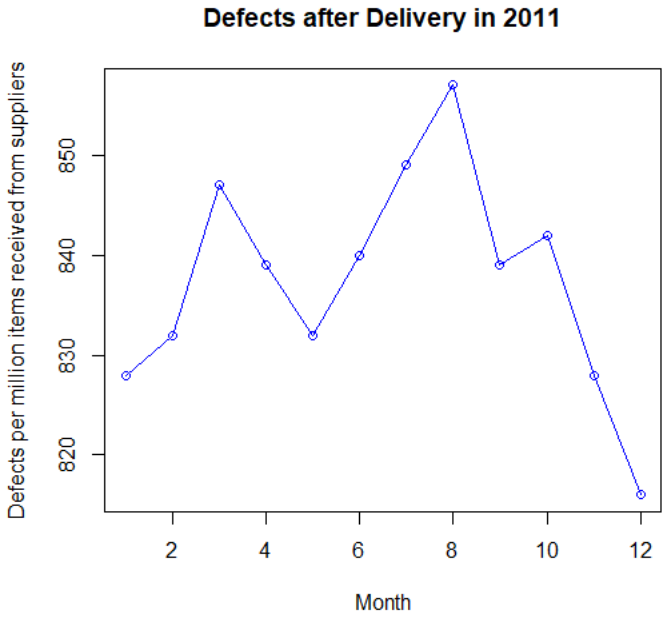


Graph 42: Line graph of the number of defects after delivery, plotted against the month the data was collected in. Each data series (each line) represents a year in the 5 years from 2010 to 2014 from which the data was collected in.

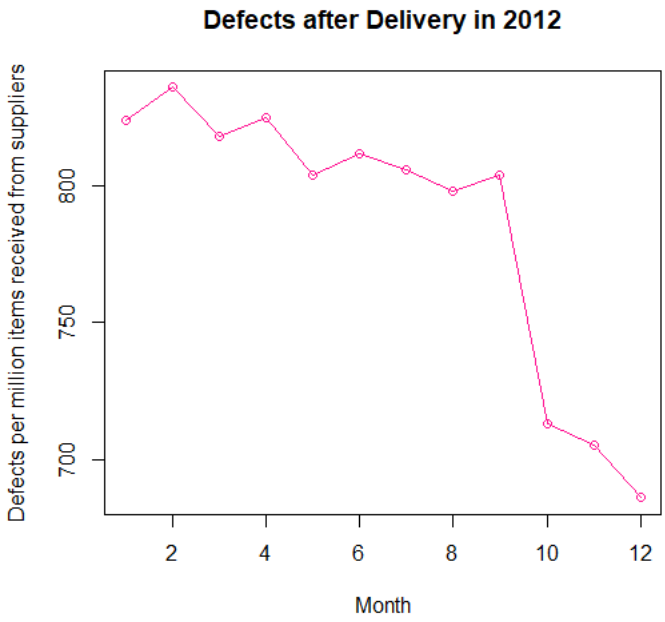
Individual line charts for each data series:



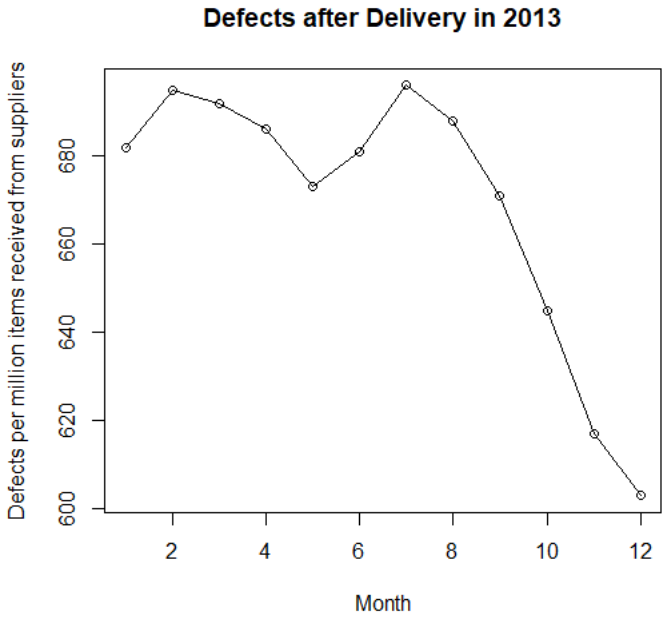
Graph 43: Line graph of the number of defects after delivery, plotted against the month in the year 2010 the data was collected in.



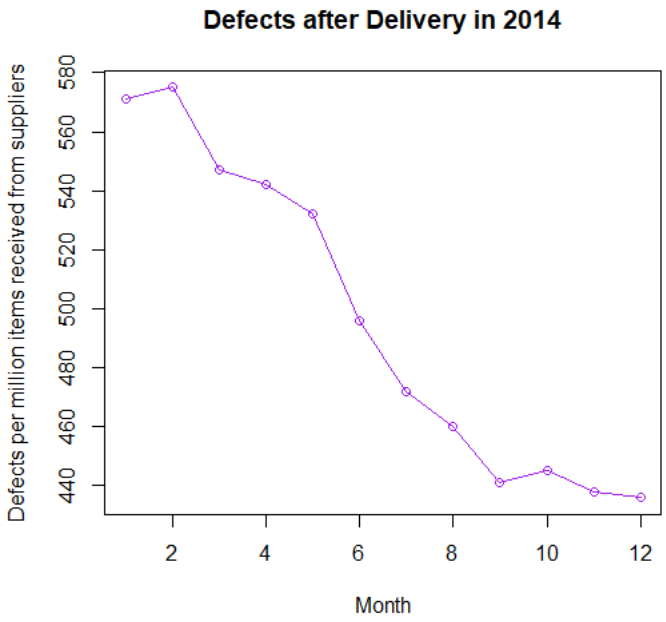
Graph 44: Line graph of the number of defects after delivery, plotted against the month in the year 2011 the data was collected in.



Graph 45: Line graph of the number of defects after delivery, plotted against the month in the year 2012 the data was collected in.



Graph 46: Line graph of the number of defects after delivery, plotted against the month in the year 2013 the data was collected in.



Graph 47: Line graph of the number of defects after delivery, plotted against the month in the year 2014 the data was collected in.

Has the number of defects after delivery per year remain the same over the past 5 years (any year that is significantly different)?

In order to analyse whether the number of defects after delivery per year remained the same over the past 5 years, we have to conduct the Bartlett test and the ANOVA test.

The Bartlett test tests for the homogeneity of variances between different groups (in this case different years).

H0: Variances in each of the groups are the same, any differences are insignificant.
H1: Variances in each of the groups are different, differences between the variances are significant.

Test result:

Bartlett test of homogeneity of variances

data: Defects by Year
Bartlett's K-squared = 38.994, df = 4, p-value = 6.985e-08

Since p-value is 6.985e-08, which is less than 0.05, reject the null hypothesis. We have sufficient evidence to support that the variances of the different groups (in this case data collected in different years) are significantly different. Differences between the variances are significant.

Although the variances in each of the groups are significantly different, the sample sizes in this case are equal in each of the groups, and hence ANOVA test should still be conducted. (i.e. because if sample sizes are equal, violation of the third assumption – homogeneity of variances between groups – does not have serious impacts.)

The ANOVA test is used to compare the means of 2 or more population groups.

H0: Mean number of defects per delivery in one of the five years (2010, 2011, 2012, 2013, 2014) is not different from the mean number of defects per delivery in any of the other four of the five years. Any difference observed is insignificant.
H1: Mean number of defects per delivery at least one of the five years (2010, 2011, 2012, 2013, 2014) is different from the mean number of defects per delivery in the other four of the five years. Differences observed are significant.

Test result:

```
            Df Sum Sq Mean Sq F value Pr(>F)
Year         4 984600  246150   178.2 <2e-16 ***
Residuals   55  75966    1381
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p-value from the ANOVA test is <2e-16, which is less than 0.05, and thus we reject the null hypothesis that the number of defects per delivery is the same across the five years, and conclude that there are significant differences in the number of defects per delivery across the 5 years.