

**KULLIYAH OF INFORMATION AND COMMUNICATION TECHNOLOGY**

**CSC 1706 PROBABILITY AND STATISTICS**

**SEMESTER 2, 2017/2018**

**SECTION 4**

**GROUP 4**

**DATA: UNEMPLOYMENT RATE STATISTICS IN MALAYSIA IN YEAR 2011, 2012, 2013, 2014, 2015 AND 2016**

**PREPARED BY:**

|  |  |
| --- | --- |
| **NAME** | **MATRIC NO.** |
| Ahmad Zaidan Bin Adnan | 1718733 |
| Muhammad Zaim Taqiyuddin Bin Zubir | 1714113 |
| Safiyyah Binti Ghazali Othman Taufik | 1718532 |
| Ahmad Zahin Fikri Bin Rozlee | 1713089 |
| Shafiq Azhar Bin Shahrizal | 1616313 |

**LECTURER**

DR. AFIDALINA TUMIAN

**DUE**

11 MAY 2018

**Introduction**

In this modern era, where many individuals and graduating students are interested with getting their job experience, some may not even have it their own way. As unpleasant as it seems, we do not always get what we want in life and hence, some are struggling in this particular problem. One of the problems is unemployment of jobs. Unfortunately, unemployment rate has been one of the main issues that need to be dealt. This issue is not only bad but also common in most countries, where people need jobs to sustain their lives, for their family and not employing them can result in a more critical result.

For the sake of making a living, if unemployment persists for quite some time, desperate individuals will look for an alternative or even worse, may even fall for performing criminal acts such as stealing, robbery and much more. To prevent this from happening in Malaysia, we must study the statistical data that have been gathered and improve on which area that requires attention. Recently, our group have been analysing the unemployment rate in each state for the past couple of years and we have organized the results in many different ways to see.

The data that our group has chosen to analyse is regarding the unemployment rate categorized by state in the years 2011, 2012, 2013, 2014, 2015 and 2016 by the Department of Statistics Malaysia. The data shows the total number of people that are unemployed in each state in Malaysia. The type of our data is population since we are studying the number of people. The type of statistics involved is descriptive statistics which consists of collection, organization, summarization and presentation of data. The type of data is quantitative discrete since our data can be ranked and there should not be any fractions in the data.

**Method**

We have constructed a line graph that shows the change of unemployment rate per year (in percentage) of each state in Malaysia. We also made a bar graph that shows the difference of the unemployment rate per state. In addition, we also managed to construct a pie chart by sorting the unemployment rate into sectors corresponding to its percentage in the data set.

After constructing the graphs, we measured the central tendency of the data such as mean, median and mode. We found the mean by adding the values of the data according to year and divided the total values of data by 13. In order to find the median, we arranged the data values according to year in ascending order and took the median of the data. We found the mode for each year by identifying the state with the highest percentage in each year. We managed to calculate the midrange of the data by observing the highest and lowest values in each year.

Other than that, we have analysed the data sets through measures of variation, which include calculating the range, variance, standard deviation and coefficient of variation. We calculated the range for each year by subtracting the highest and lowest values of the data by year. By calculating the coefficient of variation for each year, we could conclude which data (year) that is dispersed evenly.

Furthermore, we have identified the position of data using various measures of position such as quartiles, outliers and constructed the boxplot of each year from 2011 to 2016. We found the quartiles by dividing the distribution into four groups. We have successfully identified the outliers in our data according to year. Lastly, we managed to construct the boxplots for each year by using five-number summary.

**Results**

**Unemployment Rate by State in 2016**



**Unemployment Rate by State in 2015**



**Unemployment Rate by State in 2014**



**Unemployment Rate by State in 2013**



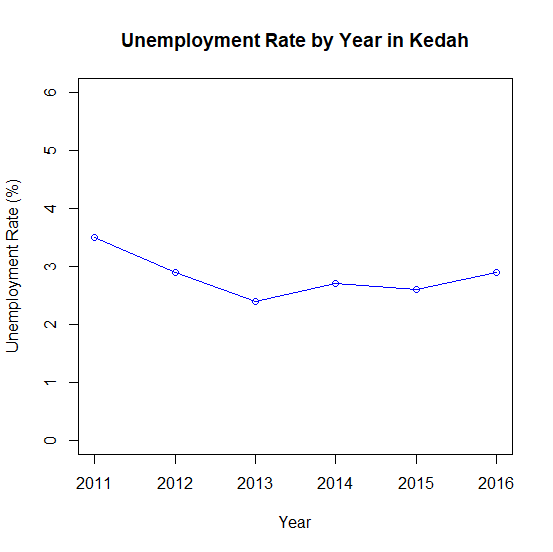
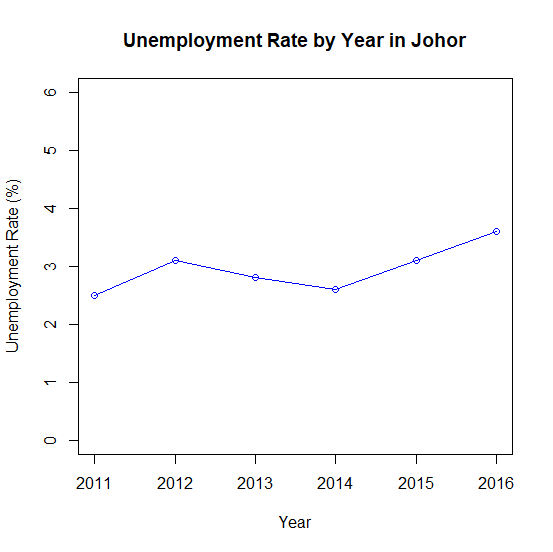
**Unemployment Rate by State in 2012**

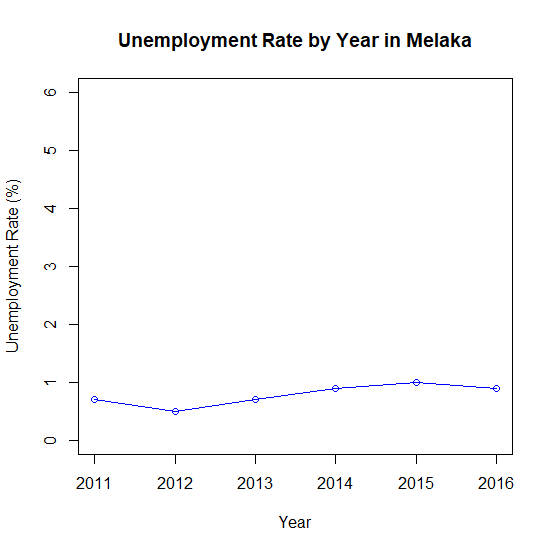
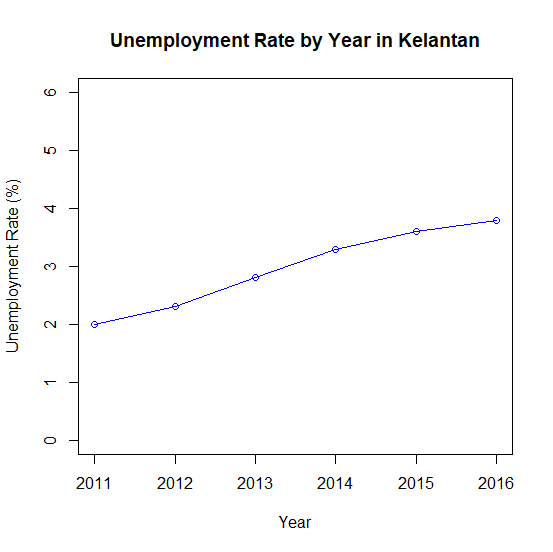


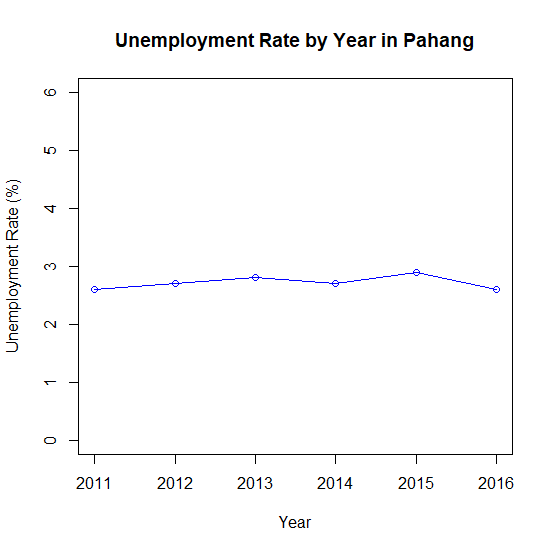
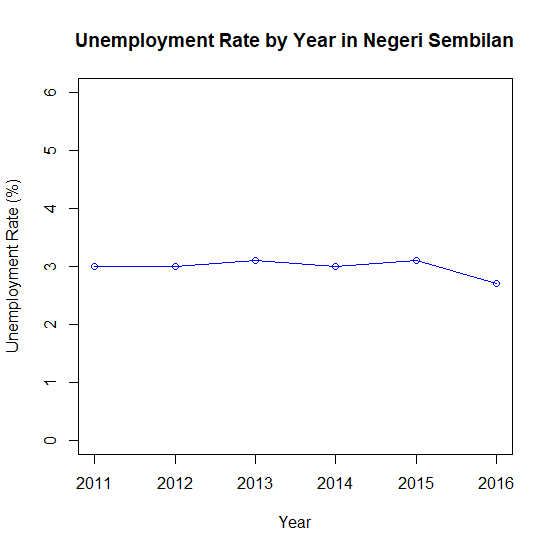
**Unemployment Rate by State in 2011**

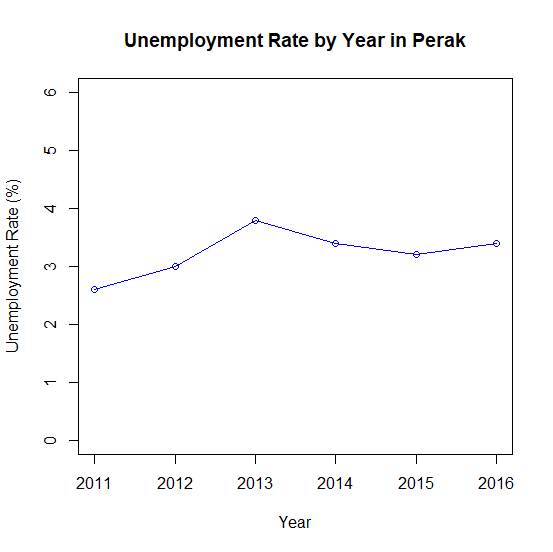
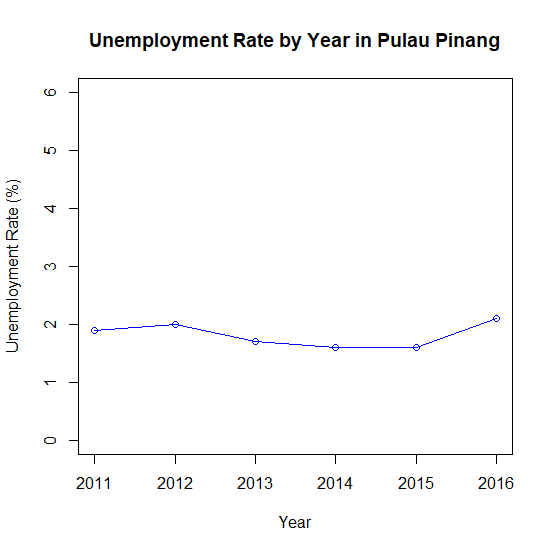


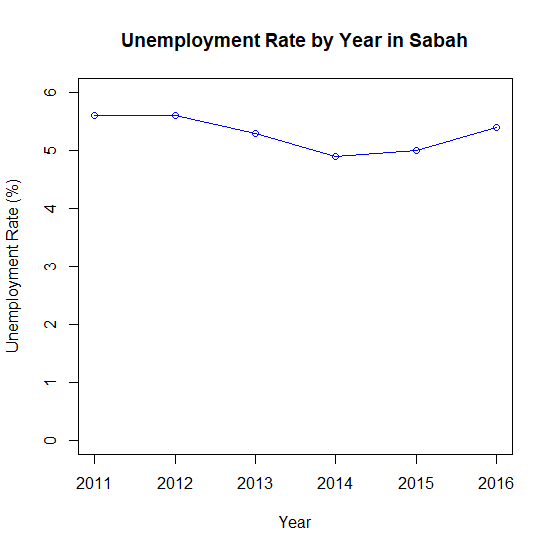
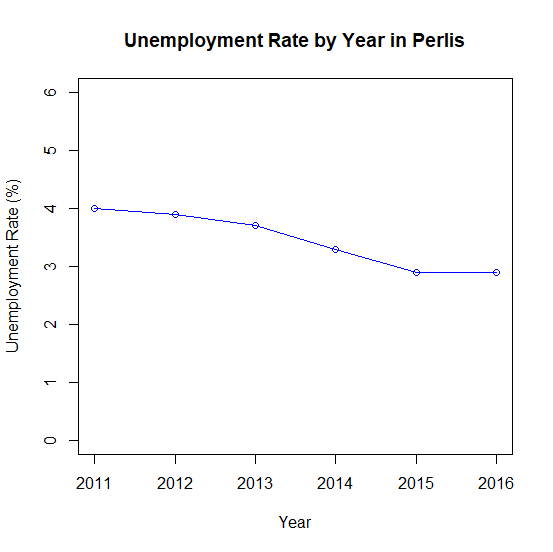
**Line Graph**

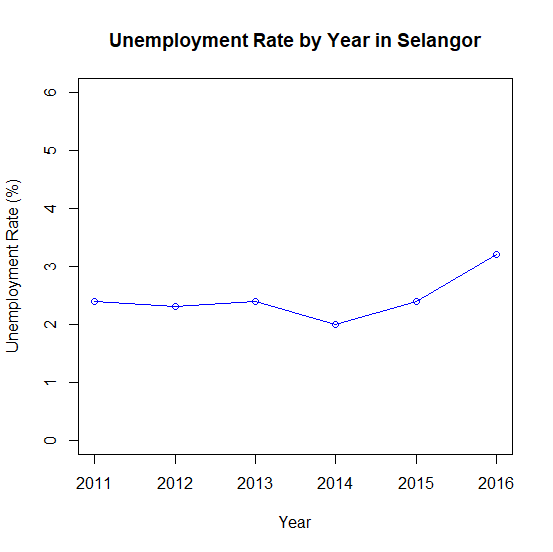
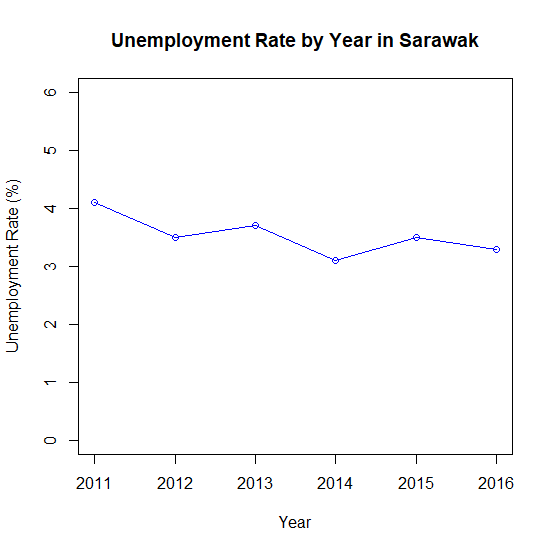


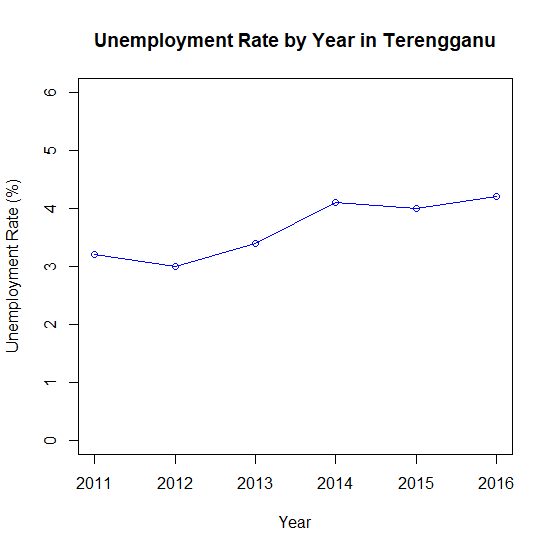












**Bar Chart**

**Legend**

**J – Johor**

**K - Kedah**

**D – Kelantan**

**M – Melaka**

**N – Negeri Sembilan**

**C – Pahang**

**A – Perak**

**R – Perlis**

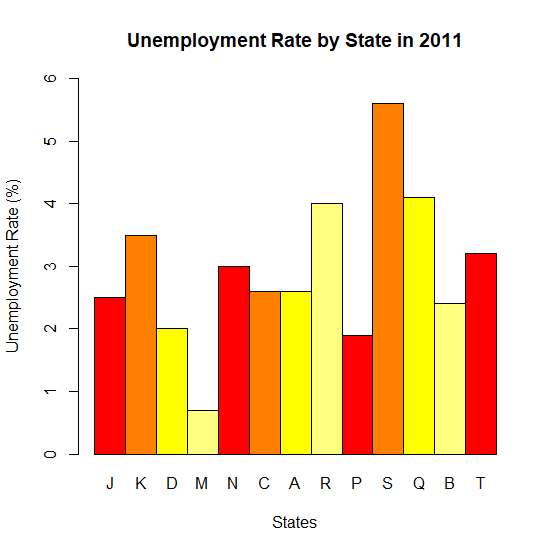
**P – Pulau Pinang**

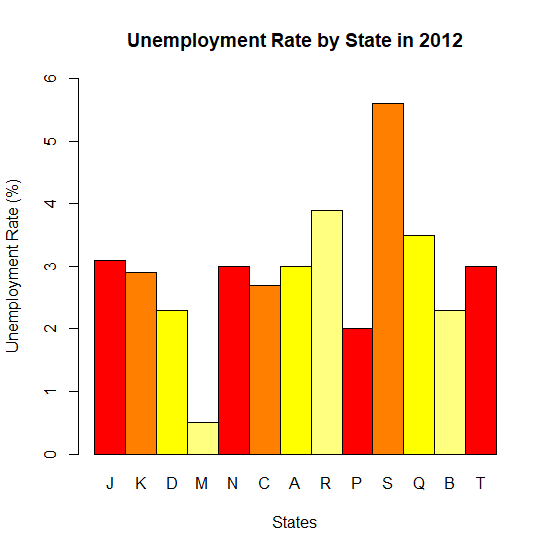
**S – Sabah**

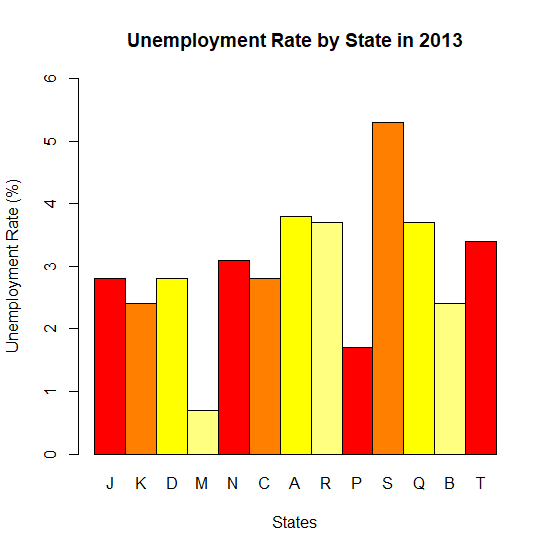
**Q – Sarawak**

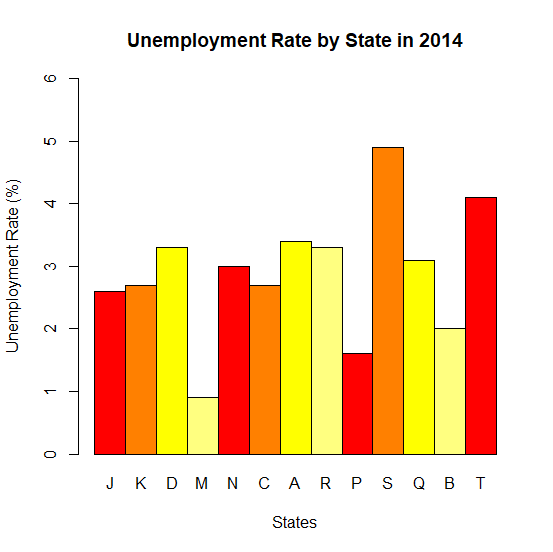
**B – Selangor**

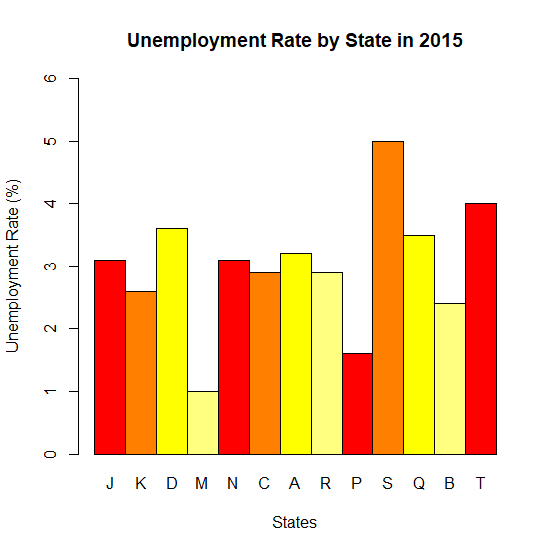
**T - Terengganu**

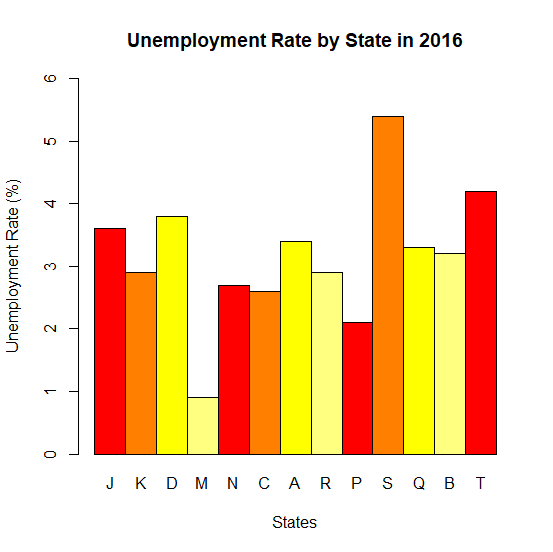












**Pie Chart**

**Legend**

**J – Johor**

**K - Kedah**

**D – Kelantan**

**M – Melaka**

**N – Negeri Sembilan**

**C – Pahang**

**A – Perak**

**R – Perlis**

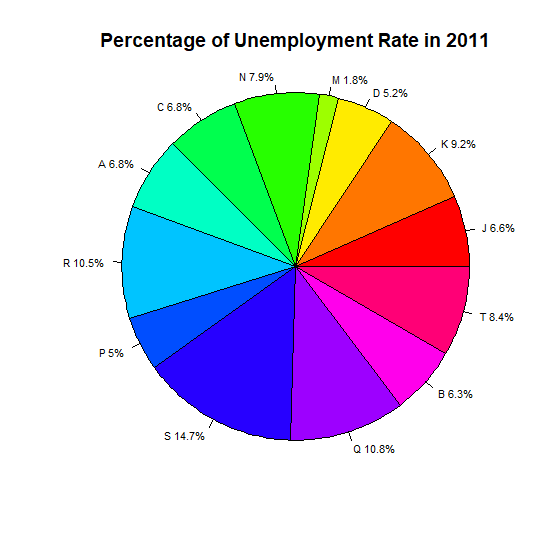
**P – Pulau Pinang**

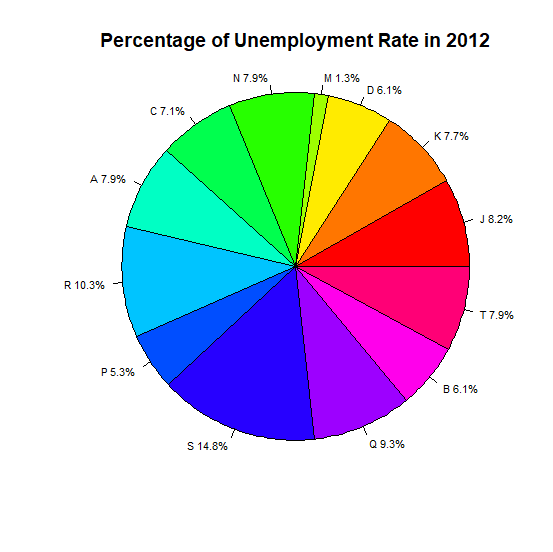
**S – Sabah**

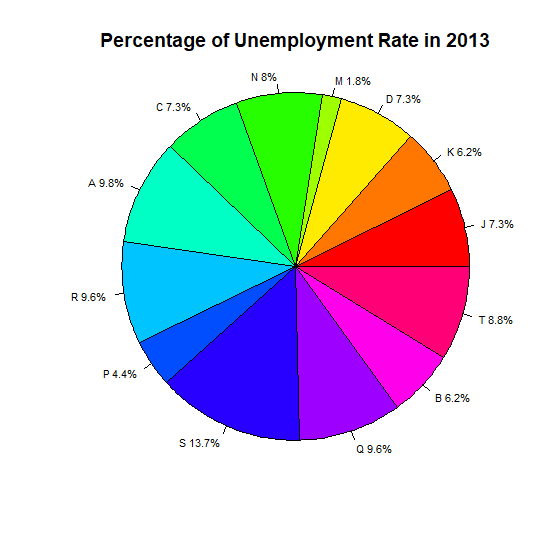
**Q – Sarawak**

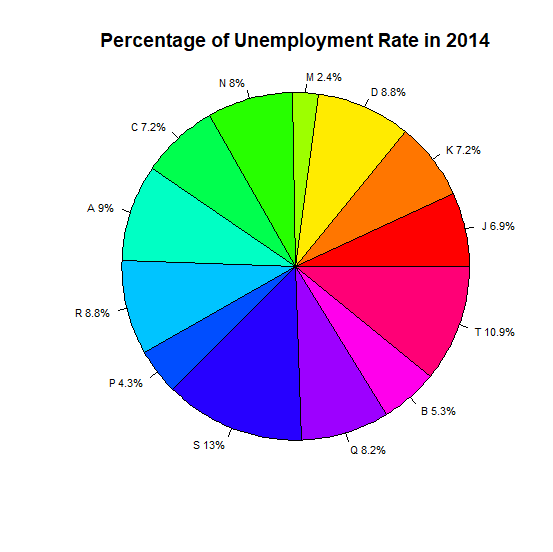
**B – Selangor**

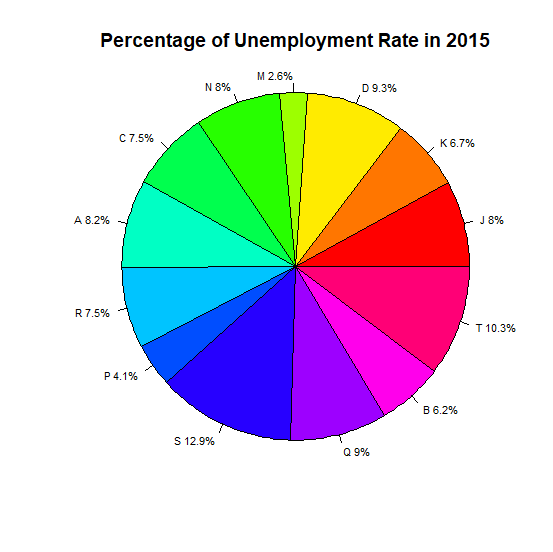
**T - Terengganu**

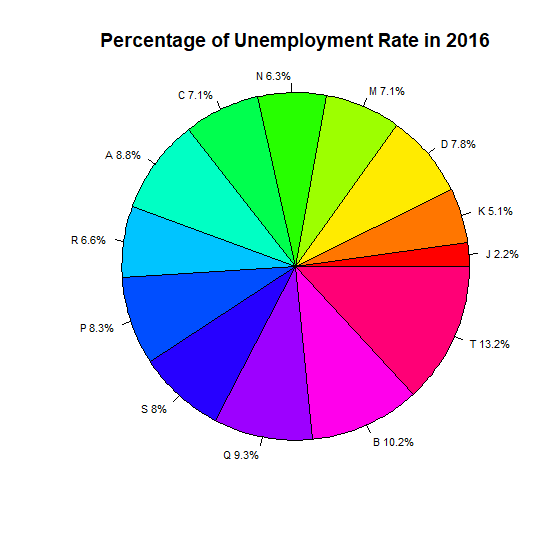




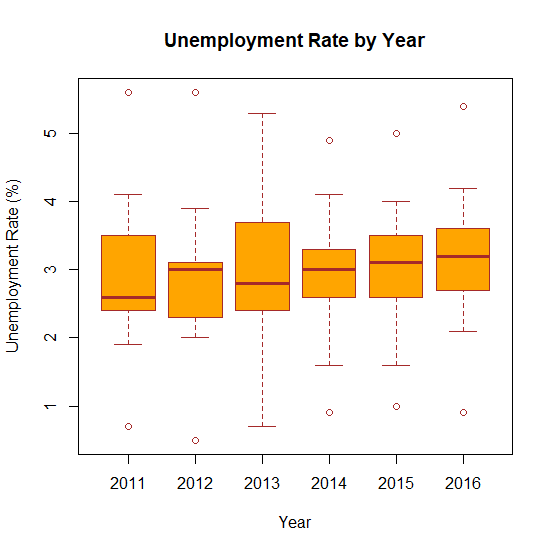








**Boxplot**

****

**Summary**



**Conclusion and Discussion**

Based on the calculation on the data, we can conclude that the mean for each years is maintain around 2.9 percent except in 2016 which is escalated slightly from 2.9923 percent in 2015 to 3.1538 percent in 2016. The lowest mean that we recorded is in 2014 at 2.8923 percent. While in 2016 the data shows the highest mean among the years which is 3.1538 percent. The median of the data is inconsistent throughout the years. In 2011, the median of the data is 2.6000 percent and increased marginally in 2012 at 3.000 percent but it decline in 2013 at 2.8000 percent. In the next years, the median of the data increase consistently from 3.000 percent to 3.2000 percent.

The variance of the data is decreasing from 2011 to 2015. In 2011 the data indicates the highest variance among the years and drop swiftly in 2012 at 1.2346 percent. The variance continues falling steadily in 2013 until 2015 and recorded the lowest value of variance in 2015 at 0.9361 percent. In contrast, the variance of the data in 2016 rise swiftly from 0.9361 percent to 1.0517 percent. The standard deviation of the data got the same pattern as the variance which falls consistently from 2011 to 2015 at 1.1645 percent to 0.9675 percent and rise slightly in 2016 from 0.9675 to 1.0255.

The minimum percentage of the data among the years stated is fluctuated and inconsistent. The minimum percentage in 2011 is 0.7 percent and fall significantly in 2012 at 0.500 percent and rise back in 2013 and 2014 at 0.7 percent and 0.9 percent respectively. In 2016 the minimum percentage of the data is decreasing at 0.900 percent. Meanwhile, the maximum percentage. While maximum value is consistent in 2011 and 2012 at 5.6 percent and fall slightly in 2013 and 2014 at 5.3 percent and 4.9 percent respectively. In 2015 and 2016 the maximum value rise gradually at 5.0 percent and 5.4 percent.

**Reference**

* http://www.data.gov.my/

**Appendix**

**Line Graph**

plot(as.matrix(data[1:6,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Johor", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[1:6,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[7:12,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Kedah", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[7:12,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[13:18,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Kelantan", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[13:18,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[19:24,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Melaka", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[19:24,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[25:30,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Negeri Sembilan", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[25:30,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[31:36,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Pahang", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[31:36,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[37:42,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Perak", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[37:42,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[43:48,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Perlis", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[43:48,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[49:54,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Pulau Pinang", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[49:54,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[55:60,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Sabah", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[55:60,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[61:66,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Sarawak", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[61:66,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[67:72,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Selangor", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[67:72,3])~as.matrix(data[1:6,2]), col="blue")

plot(as.matrix(data[73:78,3])~as.matrix(data[1:6,2]), main="Unemployment Rate by Year in Terengganu", xlab="Year", ylab="Unemployment Rate (%)", xlim=c(2011, 2016), ylim=c(0, 6), col="blue")

lines(as.matrix(data[73:78,3])~as.matrix(data[1:6,2]), col="blue")

**Bar Chart**

barplot(as.matrix(data[1:13,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2016", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

barplot(as.matrix(data[14:26,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2015", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

barplot(as.matrix(data[27:39,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2014", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

barplot(as.matrix(data[40:52,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2013", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

barplot(as.matrix(data[53:65,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2012", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

barplot(as.matrix(data[66:78,3]), beside=TRUE, names.arg=c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T"), main="Unemployment Rate by State in 2011", xlab="States", ylab="Unemployment Rate (%)", ylim=c(0,6), col=heat.colors(4), border="black")

**Pie Chart**

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2016")

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2015")

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2014")

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2013")

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2012")

pielabels<-c("J", "K", "D", "M", "N", "C", "A", "R", "P", "S", "Q", "B", "T")

pct<-round(as.matrix(data[1:13,3])/sum(as.matrix(data[1:13,3]))\*100,1)

pielabels<-paste(pielabels, pct)

pielabels<-paste(pielabels, "%", sep="")

pie(as.matrix(data[1:13, 3]), labels=pielabels, edges=300, radius=1.0, angle=45, cex=0.7, col=rainbow(length(pielabels)), main="Percentage of Unemployment Rate in 2011")

**Boxplot**

boxplot(Rate~Year,data=data,main="Unemployment Rate by Year",xlab="Year",ylab="Unemployment Rate (%)",col="orange",border="brown")

**Mean**

mean (as.matrix (data[1:13,3])

mean (as.matrix (data[14:26,3])

mean (as.matrix (data[27:39,3])

mean (as.matrix (data[40:52,3])

mean(as.matrix (data[53:65,3])

mean (as.matrix (data[66:78,3])

**Standard Deviation**

sd (as.matrix (data[1:13,3])

sd (as.matrix (data[14:26,3])

sd (as.matrix (data[27:39,3])

sd (as.matrix (data[40:52,3])

sd(as.matrix (data[53:65,3])

sd (as.matrix (data[66:78,3])

**Variance**

var (as.matrix (data[1:13,3])

var (as.matrix (data[14:26,3])

var (as.matrix (data[27:39,3])

var (as.matrix (data[40:52,3])

var (as.matrix (data[53:65,3])

var (as.matrix (data[66:78,3])

**Total**

sum (as.matrix (data[1:13,3])

sum (as.matrix (data[14:26,3])

sum (as.matrix (data[27:39,3])

sum (as.matrix (data[40:52,3])

sum (as.matrix (data[53:65,3])

sum (as.matrix (data[66:78,3])

**Project Log**

Project title : Unemployment Rate by State

Group number : 4

Section : 4

**Group members**

|  |  |  |
| --- | --- | --- |
| Name | Role | Task |
| Ahmad Zaidan bin Adnan | Task Coordinator | Divide tasks to team members and compile them to proofread the work. |
| Ahmad Zahin Fikri bin Rozlee | Team Leader | Give ideas for the project and ensure the work progress is maintained. |
| Muhammad Zaim Taqiyuddin bin Zubir | Data Collector | Collects the data for the project. |
| Safiyyah binti Ghazali Othman Taufik | Recorder | Records any update and information for the project. |
| Shafiq Azhar bin Shahrizal | Scriber | Lead the progress of writing the report. |

**Progress**

|  |  |  |  |
| --- | --- | --- | --- |
| Week | Describe your progress here | Who did what? | Any problems/challenges you face? |
| 7 | Project Proposal | Introduction - Shafiq, Safiyyah  Table Data - Zaim  Method - Zaidan, Zahin | Finding the perfect data for our project. |
| 8 | Task Division for Project | Table Data - Zahin  Histogram Data - Zaim  Frequency Polygon Data - Shafiq  Ogive Data - Safiyyah  Method - Zaidan | Have the problem in moving the data from excel to R software. |
| 9 | Reformatting for Task Division  Update on the number of years used for data | Line Graph - Zahin  Bar Chart - Shafiq Azhar  Pie Chart - Zaim  Boxplot - Safiyyah  Summary - Zaidan | Since we just have a preview of previous projects, we reformat the task so the report would satisfy the requirements like the previous ones. |
| 10 | Same progress with week 9 | Line Graph - Zahin  Bar Chart - Shafiq Azhar  Pie Chart - Zaim  Boxplot - Safiyyah  Summary - Zaidan | We have problems regarding the R codes. |
| 11 | Task division for Full Report | Introduction - Zaim  Method - Shafiq Azhar  Conclusion & Discussion - Safiyyah & Zahin | The R codes still faces some errors but we managed to fix it. |
| 12 | Compilation of parts for report | Introduction - Zaim  Method - Shafiq Azhar  Conclusion & Discussion - Safiyyah & Zahin | This week was very busy. So, we did not manage to held a meeting at this week. |
| 13 | Video presentation | Video - Zaim, Zahin, Safiyyah, Shafiq Azhar  Voice - Zaidan | The software of video was only a trial not ful version. So, we had to record the video we made from outside. |