Mortality Statistics and Cause of Death at Duhok City from The Period (2014-2019) Using R Language Data Analytics

[1]Renas Rajab Asaad, [2]Rezbar Farsat Mustafa, [3]Shiret Ismail Hussien

[1,2,3]Department of Computer Science, Nawroz University, Duhok, Kurdistan Region of Iraq

ABSTRACT

Nowadays, the amount of giving birth is increasing by amount of death, these days there are have many death causes (diseases) , in different ways, and there's different death causes in every region, In north of Iraq there is a governorate called Duhok, population of Duhok's governorate is nearly about 1.011.585 people in 2019. Descriptive tabulations of data are presented and interpreted. including rates by cause of death rate, data shown in this article are based on death causes last five years, from Dec of 2014 till the end of 2019 period of analyzing Duhok death's dataset using R programming, to analyze Age the most death in, disease the most death caused, and month the most death in, by used R. R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing.

Keywords: R Language, Mortality Statistics, Cause of Death, Duhok City, Data Analysis.

1. Introduction

This study presents death caused data from the 2014 till 2019 period linked file , In the linked file the information from the death causes is linked to information from the sex, age death in and month the most death in, for each dead person last five years of age who died in the Duhok and around, The purpose of the linkage is to analyze which age does people die the most, and which disease does the most caused to death etc.., with R programming to analyze the data.

2. Normal Data Distribution

To see if the data are normally distributed, we used chi-squared goodness-of-fit test to see If the data are distributed Skew equals 2.271687 it measures how symmetrical the data is around the midpoint , kurtosis equals 4.067458 is indicates how quickly the data's tails of and then if the mean or medians are equals , mean is 916 , median 670.5
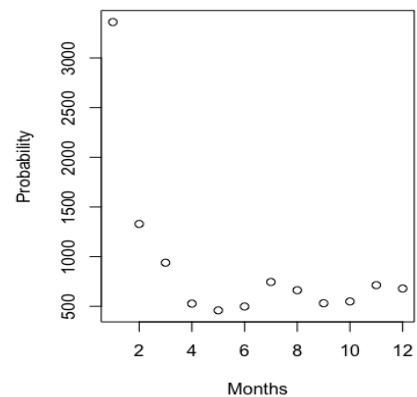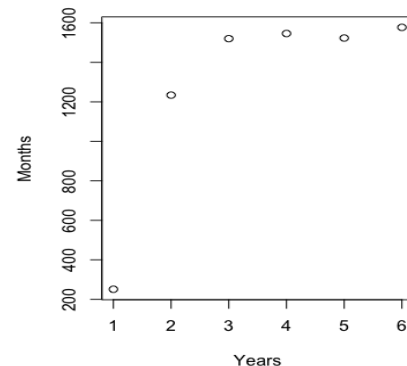

Fig. 1: Normal Data Distribution

Fig. 2: Data Distributed According to Month, Year

For years to find out if its distributed W = 0.66176, p-value = 0.002384, Skew equals -1.202573it measures how symmetrical the data is around the mid-point , kurtosis equals -0.3841762is indicates how quickly the data's tails of and then if the mean or median are equals , mean is 1275.167, median 1521.5 ,

```
//Result From R Programming Language
> V1 <- c(1,2,3,4,5)
> sum(V1)
[1] 15
> mean(V1)
[1] 3
> median(V1)
[1] 3
> range(V1)
[1] 1 5
> min(V1)
[1] 1
> max(V1)
[1] 5
> sd(V1)
[1] 1.581139
```

To find the value of correlation we have 5 variables and more than 7000 row's and the variables contain (Sex(gender) , date of birth , year of birth , month of death and death cause) to find the correlation the value must be numeric only month of death and year of death contain the numeric value's so we made the correlation between these two variables and year by year so that we can have the correlation result .

```
//Result From R Programming Language
Pearson's product-moment correlation
data: data155$YEAR and data155$MONTH
t = 6.6109, df = 1200, p-value = 5.735e-11
```

```
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1323162 0.2414397
sample estimates:     cor  0.1874563
In 2016 the Correlation between year and month of death  is
t = -14.18, df = 1579, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3791169 -0.2916319
sample estimates:     cor  -0.3360992
```

3. Correlation

Correlation for each year of all last five years, the bellow results are got from R language:

Pearson's product-moment correlation

In 2017

t = -14.18, df = 1579, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.3791169 -0.2916319

sample estimates:

cor -0.3360992

Correlation for 2017 between month of death and death year and correlation of 2017 is -0.3360992

IN 2018, the correlation is between also year and month of year

t = -13.015, df = 1581, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.3549044 -0.2658820

sample estimates:

cor -0.3110754

In 2019 Pearson's product-moment correlation data:

data199$DEATH.YEAR and data199$DEATH.YEAR

t = NA, df = 1575, p-value = NA

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

NA NA

sample estimates:

cor NA

4. Results and Discussion

After analyzing and linking data, in Duhok for last five years from Nov of 2014 till Dec of 2015, the difference between sex in death, the Male's death is 20% more than female's , and March was the month most death in , while age that has most death in was between (1930 : 1950), and disease's that caused most death where (stroke, heart attack, kidney failure , lung cancer, myocardial infraction etc..), for example the death result between last five years are shown in chart below its obvious that 2017 is 20.2% between all four years and 2014 is 3.3% and there is 0% difference between 2018 and 2016.
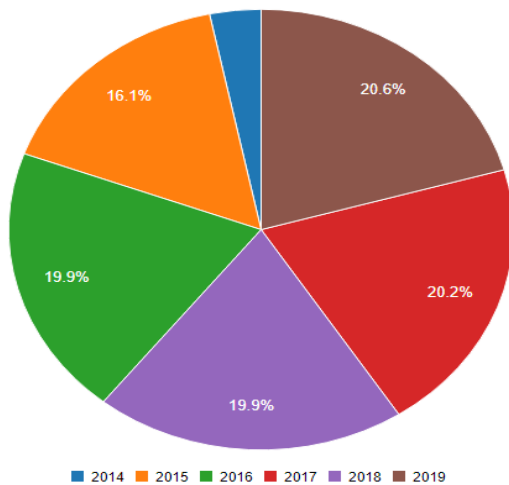

Fig. 3: death percentage from 2014 till 2019

In 2015 , the difference between sex in death, the Male's death is about 20% more than female's, and March was the month most death in, while age that has most death in was between (1943: 1998) and the most in 1950's, and disease's that caused most death where (stroke, heart attack, breast cancer, Skin cancer, Lung cancer etc..), the

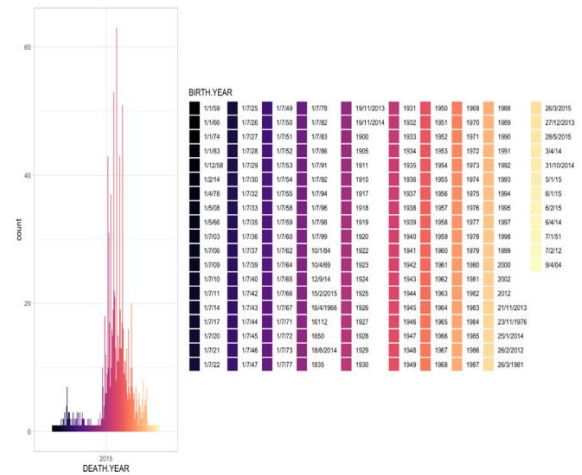chart below shows 1950 is the age most death in it.

Fig. 4: The Proportion of the dead according to Age

While In 2016 , the difference between sex in death, the Male's death is about 10% more than female's , and March was the month most death in , while age that has most death in was between (1924 : 1958) and the most in 1940's, and disease's that caused most death where (stroke, heart attack, breast cancer, bacterial blood poisoning, etc..) the chart below shows the different between month from (2014:2019) that Apr has a highest percentage and females in last five years together
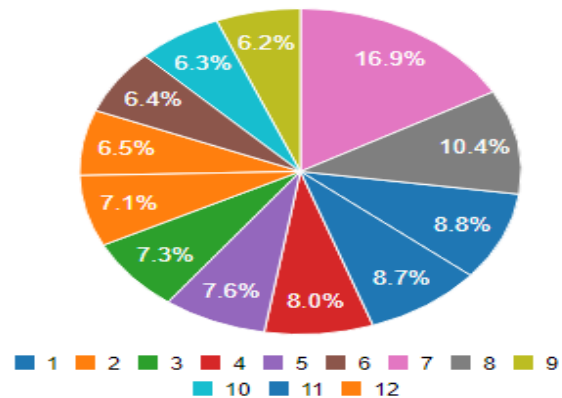

Fig. 5: Difference between death in Months from (2014:2019)

And in 2017 , the difference between sex in death, the Male's death is about 10.2% more than female's , and December was the month most death in , while age that has most death in was between (1940:1960) and the most in 1947, and disease's that caused most death where

3

(stroke, heart attack, breast cancer, heart failure, pneumonia , lung cancer, kidney failure , etc..) the chart below shows the difference between the heart attack and stroke that causes to death in 2017, it seems that stroke is increasing.
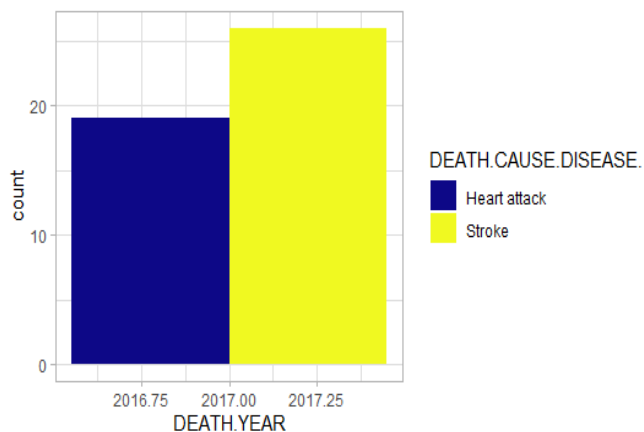

Fig. 6: difference between highest death causes (Heart attack & Stroke).

While in 2018, the difference between sex in death, the Male's death is about 10.5% more than female's , and July was the month most death in , while age that has most death in was between (1940:1960) and the most in 1947, and disease's that caused most death where (stroke, heart attack, blood cancer, heart failure, pneumonia, lung cancer, kidney failure , etc..) the chart below shows that in all last five year stroke has the highest rate, and then heart attack and kidney failure and lung cancer, and bacterial blood poisoning , cerebral hemorrhage...

```
allyears <-read.csv("allyears.csv",header = TRUE)
allyears %>%
  filter(DEATH.CAUSE.DISEASE. %in% c("death
causes")"\ %>%
  filter(DEATH.YEAR %in% "allyears") %>%
  ggplot() +
aes(x = DEATH.CAUSE.DISEASE., fill = count.) +
geom_bar(position = "dodge") +
scale_fill_viridis_d(option = "plasma") +
```
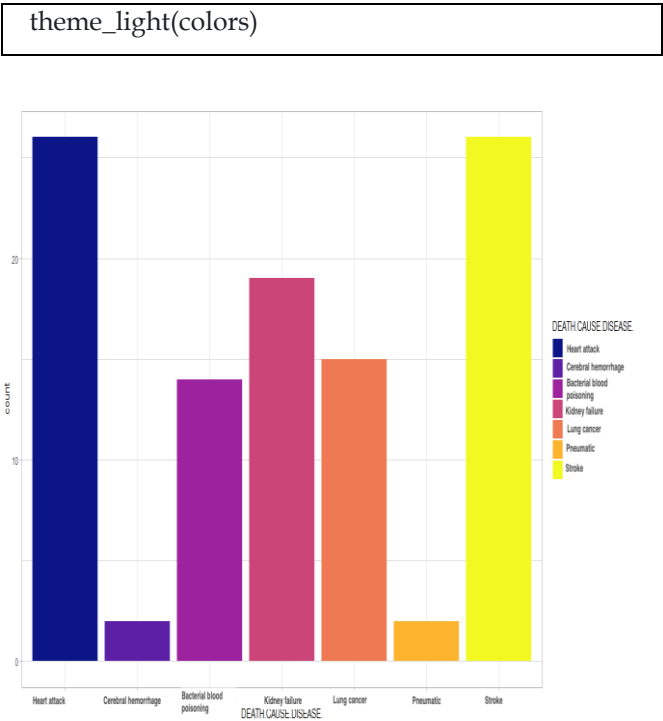
```
theme_light(colors)
```


Fig. 7: Highest death causes from (2014:2019)

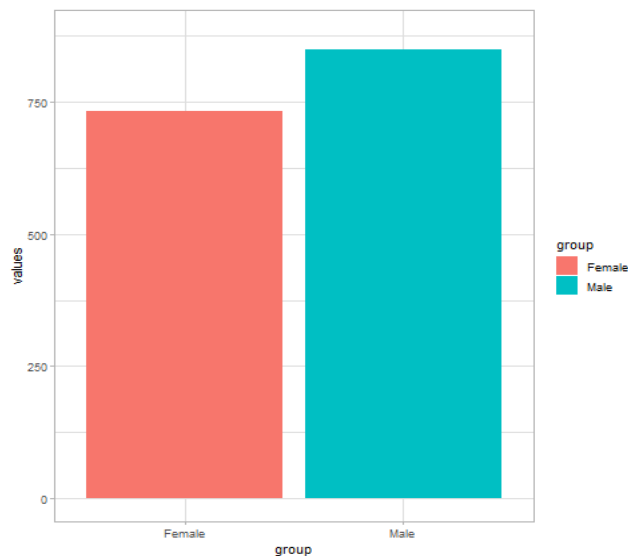This chart below shows that men's are more than female's

Fig. 8: Death Difference between male and female.

And in 2019, the difference between sex in death, the Male's death is about 9.7% more than female's , and Jan was the month most death in , while age that has most death in was between (1926:1949) and the most in 1947, and disease's that caused most death where (stroke, heart attack..) the chart below shows that in all last five year stroke has the highest rate, and then heart attack

and kidney failure and diabetic, the chart below shows Month of highest death and cause
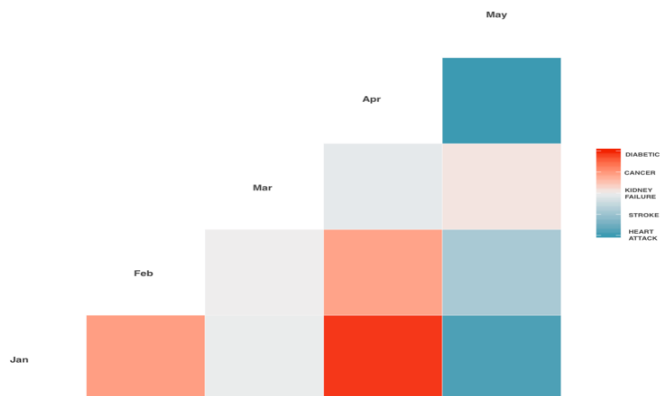

Fig. 9: Death in highest month and cause

5. Sex at death:

The sex of the most dead where man's, in Last five years, dead men's where greater than female's by 20%, the Fig. show's it accurately.
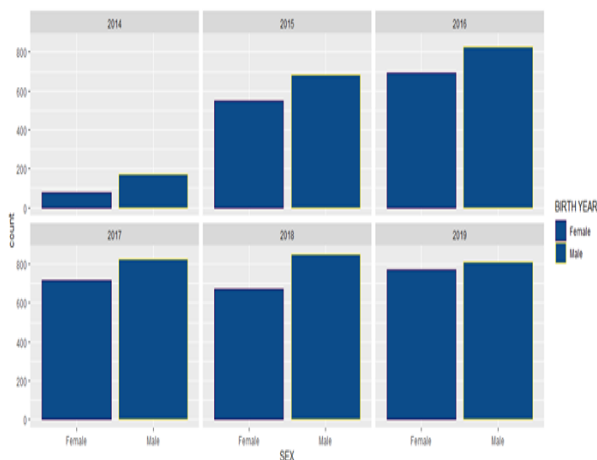

Fig. 10: Gender Death Effect in period from (2014: 2019)

6. Age at death:

Death catches every age, everyone dies in a different age, at four years old maybe or on his first month maybe, but age the most dead in was (1926: 1960) which means in 70 till 90 years old, in 2015 where between (1930 : 1950) while in 2016

- Where between (1924: 1958) but in 2017

- Where between (1935: 1960)

- While in 2018 where between (1945:1980) and form (1930:1950) in 2019

Which was the year most death in where between (1931: 1949) specially 1932's and in all four years where between 1931:1950 to the analyzed data in the chart below:
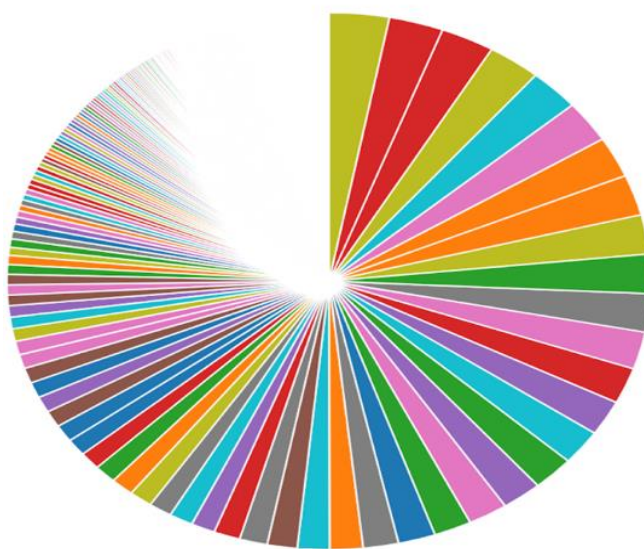

Fig. 11(A): death age chart in period (2014:2019)

Fig. 11(B): death age chart in period (2014:2019)

7. Results

according to the data's in the Duhok's main hospital descriptive tabulations of data are presented and interpreted, including rates of data by (Sex(male or female), The age most death in, Year of death, Month of death and Death cause (Disease)), after the data has been analyzed hospitals data's with R, then analyzing the difference between sex in death, the Male's death is more than female's and, in the region there are have at least more than 250 death cause's, while in every amount of time a new disease being born the most common death causes are (stroke, heart attack , kidney failure, cerebral hemorrhage, blood poisoning, lung cancer, blood cancer, breast cancer, prostate cancer, myocardial infarction, pneumonic) and the most common disease between those death causes are (stroke, heart attack, cancer and it kinds), March was the month that has most death in, in last five years most death was in 2019, while in 2015 was less than four other years , in last five years .

8. Conclusion

There are at least few or maybe 5% of human beings that live between 95 till 110 years and that every year another virus or disease are being born, and to live long or to live healthy or far from all disease that has been mentioned is to protect yourself and make tests that you need, every another month, in last five years men's where more than women's and 2019 was the year that has the most death in, compare to other five years , with R studio the data has been analyzed and had more than seven thousand people that died last five years, R studio programming language analyze the data accurately, it show's it with obvious charts and plots, finally the advice is the healthy food make people live longer.

References:

1. National Center for Health Statistics in Duhok. Public Use Data File Documentation: (2014, 2018) Period Linked Mortality data set.

2. Morandat, F., Hill, B., Osvald, L., & Vitek, J. (2012, June). Evaluating the design of the R language. In *European Conference on Object-Oriented Programming* (pp. 104-131). Springer, Berlin, Heidelberg.

3. Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717). New York: Springer.

4. Say, L., Chou, D., Gemmill, A., Tunçalp, Ö., Moller, A. B., Daniels, J., ... & Alkema, L. (2014). Global causes of maternal death: a WHO systematic analysis. *The Lancet Global Health*, 2(6), e323-e333.

5. Chamblee, R. F., & Evans, M. C. (1982). New dimensions in cause of death statistics. *American Journal of Public Health*, 72(11), 1265-1270.

6. Azadi Nation Hospital (Duhok), 2019, Public use dataset, cause of death (2014:2018) Period.

7. Lu, T. H., Lee, M. C., & Chou, M. C. (2000). Accuracy of cause-of-death coding in Taiwan: types of miscoding and effects on mortality statistics. International journal of epidemiology, 29(2), 336-343.

8. Percy, C., Stanek 3rd, E., & Gloeckler, L. (1981). Accuracy of cancer death certificates and its effect on cancer mortality statistics. American journal of public health, 71(3), 242-250.

9. Lahti, R. A., & Penttilä, A. (2003). Cause-of-death query in validation of death certification by expert panel; effects on mortality statistics in Finland, 1995. Forensic science international, 131(2-3), 113-124.

10. Team, R. C. (2000). R language definition. Vienna, Austria: R foundation for statistical computing.