Effects of Age and Gender on Assault Death Rates in Pima County vs Maricopa County

by: Sarah Stallman and Jennifer Karl

Introduction: Our goals for this project are to look at data regarding deaths in Pima County and Maricopa County. We want to see if characteristics like age and gender affect someone's susceptibility to die by any source of violence or assault and if our results will be similar for both counties. We also want to find out which county has a higher assault death rate.

We're specifically looking at Pima County and Maricopa County because Pima is currently the county we are in at the University of Arizona and Maricopa is the county that Phoenix is in which is where a lot of people who go to this school are from. It is useful to recognize the potential threats in our community and to see what may possibly impact the likelihood of being a victim of assault. Since these are the two most populated counties in Arizona, we want to see if the trends will be consistent in both counties.

We have been able to find a variety of different studies that look at death rates in Arizona for various causes such as child drowning rates compared to neglect and murder rates in Phoenix compared yearly but have not come across any studies that compare age and gender to assault death rates while simultaneously comparing those findings between the two most populated counties in Arizona.

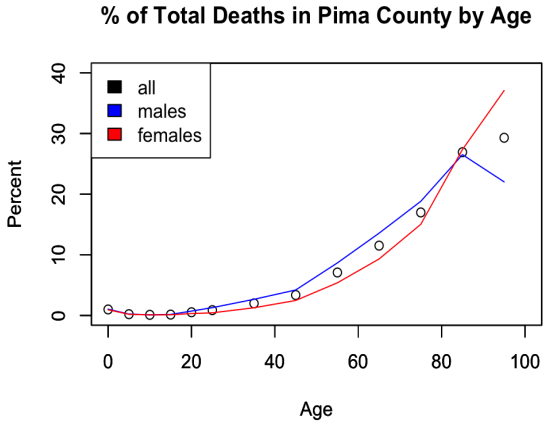
Methods: Our data displays the amount of deaths per age for all causes in Pima County and Maricopa County between 1999-2015 as well as the amount of deaths per age for assault in these two counties for the same time period.

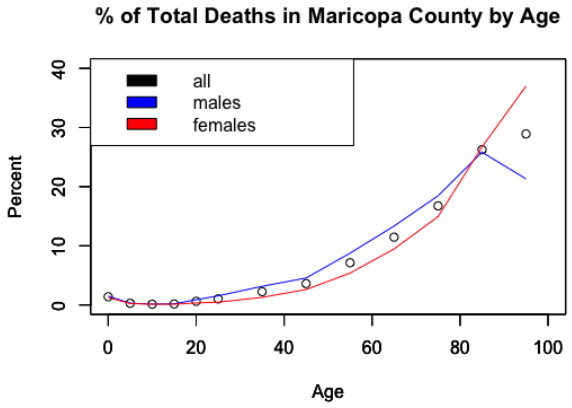
According to the CDC webpage this data is collected through all death certificates in every state as well as the District of Columbia, however for this project we are only looking at deaths in Pima County and Maricopa County. Directly from the CDC webpage: "Deaths of nonresidents (e.g. nonresident aliens, nationals living abroad, residents of Puerto Rico, Guam, the Virgin Islands, and other territories of the U.S.) and fetal deaths are excluded. Mortality data from the death certificates are coded by the states and provided to NCHS through the Vital Statistics Cooperative Program or coded by NCHS from copies of the original death certificates provided to NCHS by the State registration offices."

We believe there is little to no potential for any bias in this data set due to the fact that every single death is recorded.

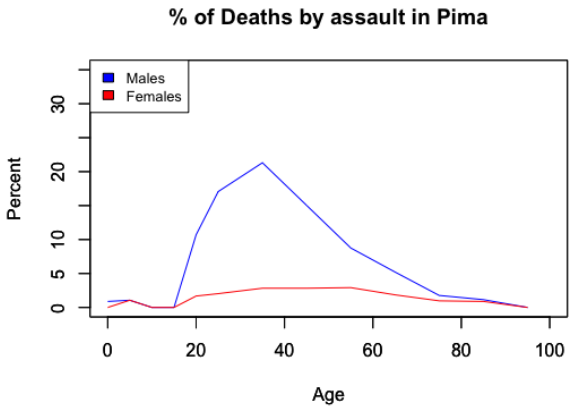
Results: As we can see in the graphs below, there is nearly an exponential increase in the number of deaths by age due to all causes of death for both males and females as age increases in both Pima and Maricopa counties. This is expected because people, both male and female, have a higher chance of dying as they get older. However, we can see that the data for assault deaths does not follow the same trend. When looking at the data for all deaths we saw that the risk of death increased with age whereas when looking at the data for assault deaths we see that people ranging in age 20-40 actually have the highest susceptibility to death by assault and people in the highest age range, 80-100, have a fairly low susceptibility. Again, this is the same for both of the counties we are looking at.

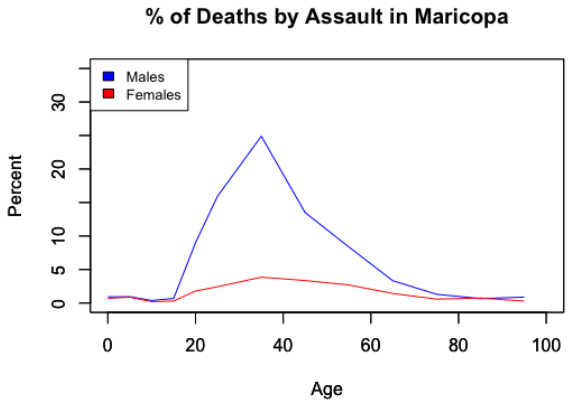
These graphs display the percent of deaths by age for each gender. The trend between male and female death rates for all causes is relatively similar. For both genders in both counties an increase in age corresponds to an increase in the likelihood of death by any cause.



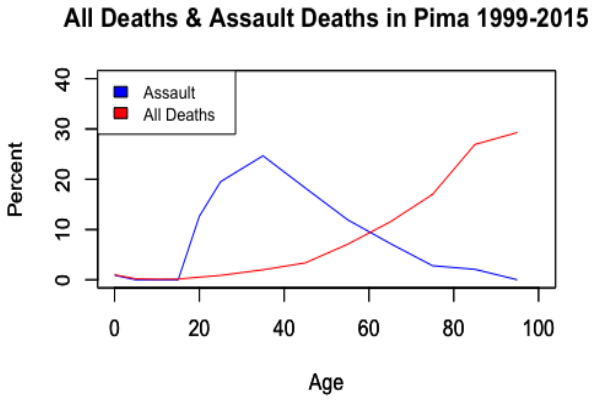


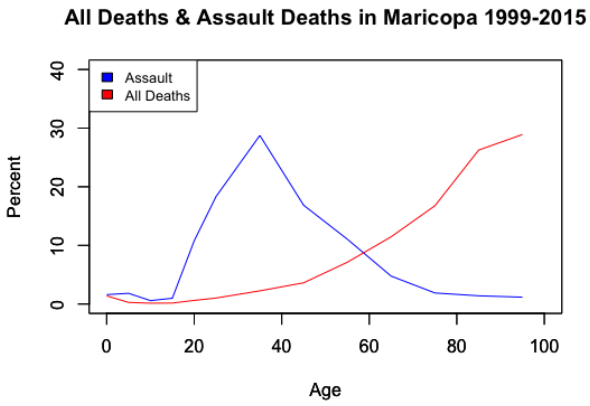
These graphs display the percent of assault deaths by age for each gender. The trends are consistent in both counties, however we see a spike in assault deaths for males in the 20-60 age range.



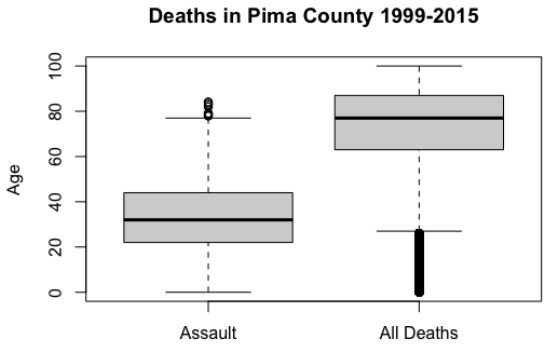


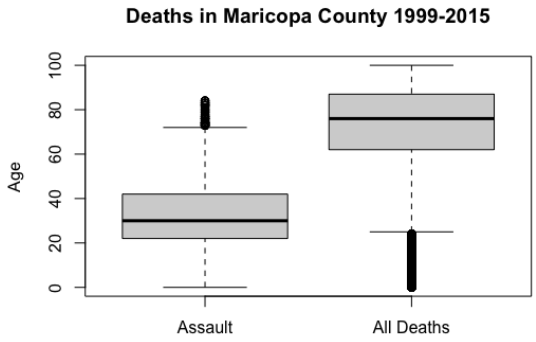
These graphs compare the percentage of deaths by age for all deaths and assault deaths. In both counties, older people have a higher likelihood of dying in general but the amount of assault deaths dominates the 20-60 age range with people in their 30s being at the highest risk.



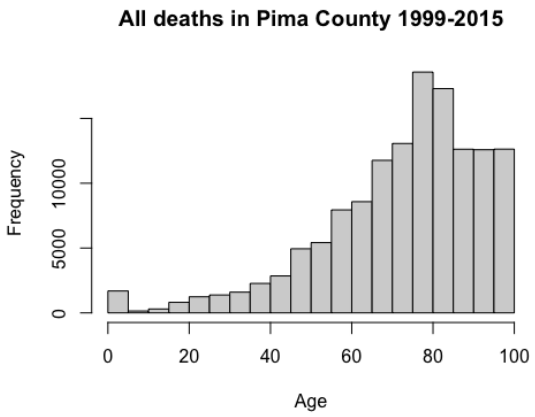


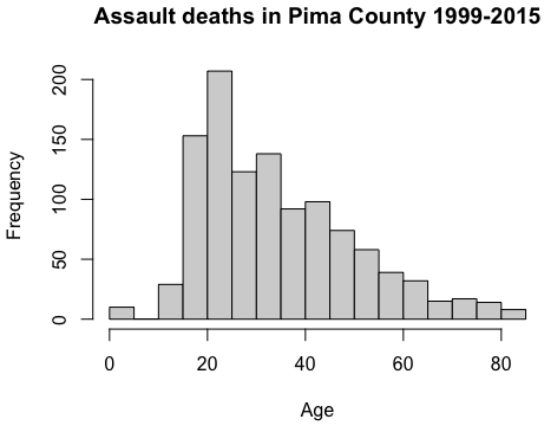
These histograms display the quartile ranges of number of deaths by age for both assault deaths and all deaths in both counties. As we can see the majority of all deaths occur for people of ages ranging from about the 65 year mark to the 90 year mark with the median falling in the late 70s. This remains the same for both Pima County and Maricopa County. However, when it comes to assault deaths the majority of people affected are between the ages of about 25 to about 45 with the median falling in the early thirties. Again, these results are nearly identical for both Pima and Maricopa. To produce these histograms, we needed age at the time of death whereas our data had the number of individuals who died from a cause in an age range, so we had R make a random sample of ages based on the number of individuals who died in each age range for all deaths and for assaults.

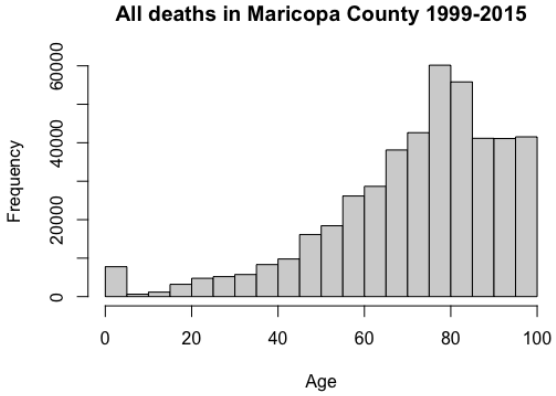


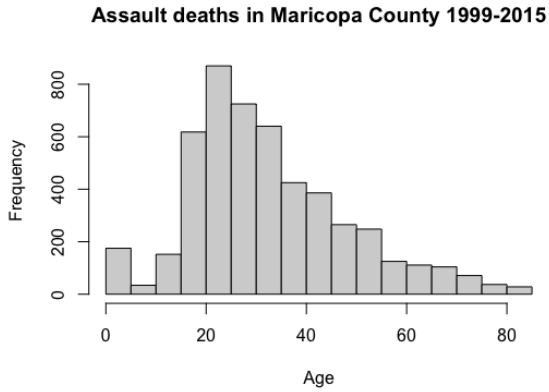


These bar graphs portray the difference between the number of total deaths and the number of assault deaths regardless of gender in both counties. Again we can see relatively the same trends in these bar graphs as we have been seeing in the rest of the graphs.









5 Number Summaries Pima County Arizona:

Age of Death from all Causes:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	63.0	77.0	72.9	87.0	100.0

Age of Death Due to Assault:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	22.00	31.00	34.62	44.00	84.00

5 Number Summaries Maricopa County Arizona:

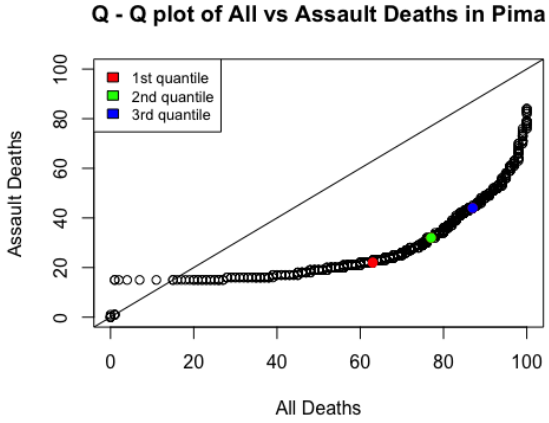
Age of Death from all Causes:

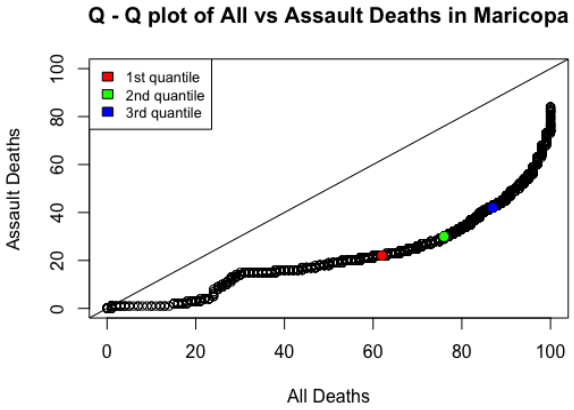
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	62.00	76.00	72.08	87.00	100.00

Age of Death Due to Assault:

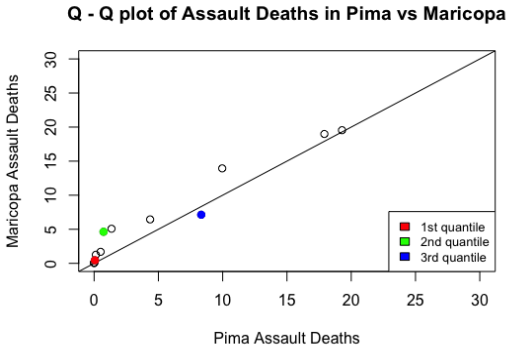
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	22.00	30.00	32.99	42.00	84.00

These Q-Q plots show that the majority of deaths are caused by things other than assault and we can also see that there is not a normal distribution.





This Q-Q plot of the assault deaths as a percent of total deaths in each county shows normal distribution and seems to mildly favor the Maricopa county side of the plot suggesting that residents of Maricopa county are more likely to die by assault than residents of Pima County. We performed a Paired T-Test to see if this observation is statistically significant.



Paired T-Test

data: The percent of people in all age groups who die of assault as opposed to all causes of death in Maricopa and Pima county in Arizona.

H0: There is no difference in the means of the percent of people who die from assault (as opposed to all causes of death) in Pima county and Maricopa county.

H1: There is a difference in the means of the percent of people who die from assault (as opposed to all causes of death). Maricopa county residents have a larger mean percentage of assault deaths than Pima county.

Paired T-Test result in R:

t = 2.0388, df = 12, p-value = 0.03207

Alternative hypothesis: true mean difference is greater than 0

95 percent confidence interval:

0.1621449 Inf

sample estimates:

mean difference

1.288917

Result: The p-value of 0.03207 is less than 0.05 and zero is not in the confidence interval. Therefore, at a 95% confidence level we reject the null hypothesis in favor of the alternative hypothesis. Therefore, the percentage of people who die from assault is greater in Maricopa county than in Pima county.

Discussion: Through our research and data analysis of death rates in Pima and Maricopa Counties we have come to many conclusions that answer our questions: Do age and gender affect someone's susceptibility to die by any source of violence or assault? Are these results consistent in both counties? And are the percentages of assault deaths in one county more than the other or are they the same?

We have concluded that young adults and middle aged people are more likely to die by assault in both counties. We have also found that young men are at the highest risk. Between the two counties we've inferred that Maricopa County has a higher rate of assault deaths than Pima County does. This poses the bigger question of: Why?

Some of our theories include the fact that Maricopa County has a much higher population than any other county in Arizona and this is because Phoenix, a major metropolitan area, is in Maricopa County and highly populated urban cities are known to have higher crime rates. What was most surprising from our findings was that young men had the highest assault death rate. This is surprising because young men have the highest chances of being able to defend themselves over elders and women. To get to the bottom of this would require looking into many other factors such as social and economic status. Then we could try to make connections between these various factors if there were any more consistent patterns in the trends.

Overall, we were able to come to some good conclusions through our modeling of our data.

References:

Centers for Disease Control and Prevention. (2022). *Compressed mortality file on CDC Wonder*. Centers for Disease Control and Prevention. Retrieved September 20, 2022, from https://wonder.cdc.gov/mortsql.html

Appendix:

```
#create data for all deaths in Pima county by age group

pimadeath<-c(1365,288,142,194,726,1206,2740,4636,9753,15855,23393,37072,40345)


#define age groups

#*(age is in ranges <1 , 1-4, 5-9, 10-14, 15-19,20-24,

# 25-34,35-44, 45-54, 55-64, 65-74, 75-84, 85+)

x<-c(0,5,10,15,20,25,35,45,55,65,75,85,95)


#Plot Total Deaths in Pima by age

plot(x,pimadeath,

        xlab="Age of Death",

        ylab="Total Deaths",

        main = "Deaths in Pima County by age from 1999-2015",

        xlim=c(0,100),

        type = "l")


#Create new data for Total deaths by age group as a percent of the total deaths

pertotal<-(100*pimadeath/sum(pimadeath))


#Plot total deaths per age group in pima county as a percent of the total deaths

plot(x,pertotal,

        main="% of Total Deaths in Pima County by Age",

        xlab="Age",

        ylab="Percent",
```

```
                xlim=c(0,100),
                ylim = c(0,40))

#Create data Total deaths for males by age group
mtotal<-c(741,159,74,118,515,917,1896,2991,6170,9658,13397,18873,15676)
pmtotal<-c(100*mtotal/sum(mtotal))

#Create Data for total deaths for females by age group
ftotal<-c(624,129,68,76,211,289,844,1645,3583,6197,9996,18199,24669)
pftotal<-c(100*ftotal/sum(ftotal))

#Add the female and male totals to the graph
par(new=TRUE)
plot(x,pmtotal,
                xlab="Age",
                ylab="Percent",
                xlim=c(0,100),
                ylim=c(0,40),
                type="l",
                col="blue")
par(new=TRUE)
plot(x,pftotal,
                xlab="Age",
                ylab="Percent",
                xlim=c(0,100),
                ylim=c(0,40),
                type="l",
                col="red")
legend("topleft",
                c("all","males","females"),
                fill =c("black","blue","red"))
```

```
#look at assault as the cause of death

assault<-c(10,24,0,0,140,216,273,202,132,80,31,23,0)

plot(x,assault)

passault<-c(100*assault/sum(assault))

plot(x,passault,

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,40),

        type="l",

        col="blue")

par(new=TRUE)

plot(x,pertotal,

        main="All Deaths & Assault Deaths in Pima 1999-2015",

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,40),

        type="l",

        col="red")

legend("topleft",

        c("Assault","All Deaths"),

        fill = c("blue","red"),

        cex=0.8)


#Look at likelihood of death by assault if male vs female


#Assault deaths of males and females

amale<-c(10,12,0,0,121,193,241,170,99,59,20,13,0)

afemale<-c(0,12,0,0,19,23,32,32,33,21,11,10,0)
```

```
massault<-c(100*amale/sum(assault))

fassault<-c(100*afemale/sum(assault))



#Plot male vs female assault deaths

plot(x,massault,

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,35),

        type="l",

        col="blue")

par(new=TRUE)

plot(x,fassault,

        main="% of Deaths by assault in Pima",

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,35),

        type="l",

        col="red")

legend("topleft",

        c("Males","Females"),

        fill = c("blue","red"),

        cex=0.8)



#make a sample of the ages when an individual

#died from any cause in pima county (Since data was in age ranges

#we need to make a sample of the possible ages)

bages<-sample(c(0:1),1365,replace=TRUE)
```

```
bbages<-sample(c(1:4),288,replace=TRUE)

bcages<-sample(c(5:9),142,replace=TRUE)

bdages<-sample(c(10:14),194,replace=TRUE)

beages<-sample(c(15:19),726,replace=TRUE)

bfages<-sample(c(20:24),1206,replace=TRUE)

bgages<-sample(c(25:34),2740,replace=TRUE)

bhages<-sample(c(35:44),4636,replace=TRUE)

biages<-sample(c(45:54),9753,replace=TRUE)

bjages<-sample(c(55:64),15855,replace=TRUE)

bkages<-sample(c(65:74),23393,replace=TRUE)

blages<-sample(c(75:84),37072,replace=TRUE)

bmages<-sample(c(85:100),40345,replace=TRUE)

allages<-c(bages,bbages,bcages,bdages,beages,bfages,bgages,bhages,biages,bjages,bkages,blages,bmages)


#look at 5 number summary for all deaths in pima county

summary(allages)


#make a sample of the ages when an individual died of an assault

# in Pima County (Since data was in age ranges

# we need to make a sample of the possible ages)

aages<-sample(c(0:1),10,replace=TRUE)

abages<-sample(c(15:19),140,replace=TRUE)

acages<-sample(c(20:24),216,replace=TRUE)

adages<-sample(c(25:34),273,replace=TRUE)

aeages<-sample(c(35:44),202,replace=TRUE)

afages<-sample(c(45:54),132,replace=TRUE)

agages<-sample(c(55:64),80,replace=TRUE)

ahages<-sample(c(65:74),31,replace=TRUE)

aiages<-sample(c(75:84),23,replace=TRUE)

assaultages<-c(aages,abages,acages,adages,aeages,afages,agages,ahages,aiages)
```

```
#look at 5 number summary and plots for assaults in Pima

summary(assaultages)


#look at differences between all deaths and death by assault

boxplot(assaultages,allages,

        main=c("Deaths in Pima County 1999-2015"),

        ylab=c("Age"),

        names=c("Assault","All Deaths"))


hist(allages,

        main=c("All deaths in Pima County 1999-2015"),

        xlab=c("Age"),)


hist(assaultages,

        main=c("Assault deaths in Pima County 1999-2015"),

        xlab=c("Age"))


#Plot Q - Q plot of deaths

#This is not normal distribution

qqplot(sort(allages),sort(assaultages),

        xlim=c(0,100),

        ylim=c(0,100),

        col="black",

        xlab = c("All Deaths"),

        ylab = c("Assault Deaths"),

        main=c("Q - Q plot of All vs Assault Deaths in Pima"))

abline(a=0,b=1)

q<-c(0.25,0.50,0.75)

points(quantile(allages,q),quantile(assaultages,q),col=rainbow(3),pch=19)

legend("topleft",

        c("1st quantile","2nd quantile", "3rd quantile"),
```

```
                fill = c(rainbow(3)),
                cex=0.8)



#PART 2
#Look  at Maricopa County



#create data for all deaths in Maricopa county by age group
mardeath<-c(6350,1276,626,760,2828,4672,10266,16513,32562,52340,76485,119844,131957)


#define age groups
#*(age is in ranges <1 , 1-4, 5-9, 10-14, 15-19,20-24,
# 25-34,35-44, 45-54, 55-64, 65-74, 75-84, 85+)
x<-c(0,5,10,15,20,25,35,45,55,65,75,85,95)


#Plot Total Deaths in Maricopa by age
plot(x,mardeath,
                xlab="Age of Death",
                ylab="Total Deaths",
                main = "Deaths in Maricopa County by age from 1999-2015",
                xlim=c(0,100),
                type = "l")


#Create new data for Total deaths by age group as a percent of the total deaths
mpertotal<-(100*mardeath/sum(mardeath))


#Plot total deaths per age group in Maricopa county as a percent of the total deaths
plot(x,mpertotal,
                main="% of Total Deaths in Maricopa County by Age",
                xlab="Age",
```

```
                ylab="Percent",
                xlim=c(0,100),
                ylim = c(0,40))

#Create data Total deaths for males by age group
mmtotal<-c(3542,715,351,462,2089,3636,7426,10744,20650,31416,43473,60818,50196)
mpmtotal<-c(100*mmtotal/sum(mmtotal))

#Create Data for total deaths for females by age group
mftotal<-c(2808,561,275,298,739,1036,2840,5769,11912,20924,33012,59026,81761)
mpftotal<-c(100*mftotal/sum(mftotal))

#Add the female and male totals to the graph
par(new=TRUE)
plot(x,mpmtotal,
                xlab="Age",
                ylab="Percent",
                xlim=c(0,100),
                ylim=c(0,40),
                type="l",
                col="blue")
par(new=TRUE)
plot(x,mpftotal,
                xlab="Age",
                ylab="Percent",
                xlim=c(0,100),
                ylim=c(0,40),
                type="l",
                col="red")
legend("topleft",
                c("all","males","females"),
```

```
                fill =c("black","blue","red"))


#look at assault as the cause of death

marassault<-c(80,91,29,49,537,913,1431,839,550,237,94,70,58)

plot(x,marassault)

marpassault<-c(100*marassault/sum(marassault))

plot(x,marpassault,

                xlab="Age",

                ylab="Percent",

                xlim=c(0,100),

                ylim=c(0,40),

                type="l",

                col="blue")

par(new=TRUE)

plot(x,mpertotal,

                main="All Deaths & Assault Deaths in Maricopa 1999-2015",

                xlab="Age",

                ylab="Percent",

                xlim=c(0,100),

                ylim=c(0,40),

                type="l",

                col="red")

legend("topleft",

                c("Assault","All Deaths"),

                fill = c("blue","red"),

                cex=0.8)


#Look at likelihood of death by assault if male vs female


#Assault deaths of males and females

mamale<-c(46,47,19,34,448,792,1239,672,416,166,65,33,44)
```

```
mafemale<-c(34,44,10,15,89,121,192,167,134,71,29,37,14)

mmassault<-c(100*mamale/sum(marassault))

mfassault<-c(100*mafemale/sum(marassault))



#Plot male vs female assault deaths

plot(x,mmassault,

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,35),

        type="l",

        col="blue")

par(new=TRUE)

plot(x,mfassault,

        main="% of Deaths by Assault in Maricopa",

        xlab="Age",

        ylab="Percent",

        xlim=c(0,100),

        ylim=c(0,35),

        type="l",

        col="red")

legend("topleft",

        c("Males","Females"),

        fill = c("blue","red"),

        cex=0.8)



#make a sample of the ages when an individual

#died from any cause in Maricopa county (Since data was in age ranges

#we need to make a sample of the possible ages)
```

```
mbages<-sample(c(0:1),6350,replace=TRUE)

mbbages<-sample(c(1:4),1276,replace=TRUE)

mbcages<-sample(c(5:9),626,replace=TRUE)

mbdages<-sample(c(10:14),760,replace=TRUE)

mbeages<-sample(c(15:19),2828,replace=TRUE)

mbfages<-sample(c(20:24),4672,replace=TRUE)

mbgages<-sample(c(25:34),10266,replace=TRUE)

mbhages<-sample(c(35:44),16513,replace=TRUE)

mbiages<-sample(c(45:54),32562,replace=TRUE)

mbjages<-sample(c(55:64),52340,replace=TRUE)

mbkages<-sample(c(65:74),76485,replace=TRUE)

mblages<-sample(c(75:84),119844,replace=TRUE)

mbmages<-sample(c(85:100),131957,replace=TRUE)

mallages<-c(mbages,mbbages,mbcages,mbdages,mbeages,mbfages,mbgages,mbhages,mbiages,mbjages,mbkages,mblages,mbmages)


#look at 5 number summary for all deaths in pima county

summary(mallages)


#make a sample of the ages when an individual died of an assault

# in Maricopa County (Since data was in age ranges

# we need to make a sample of the possible ages)

maages<-sample(c(0:1),80,replace=TRUE)

mabages<-sample(c(1:4),91,replace=TRUE)

macages<-sample(c(5:9),29,replace=TRUE)

madages<-sample(c(10:14),49,replace=TRUE)

maeages<-sample(c(15:19),537,replace=TRUE)

mafages<-sample(c(20:24),913,replace=TRUE)

magages<-sample(c(25:34),1431,replace=TRUE)

mahages<-sample(c(35:44),839,replace=TRUE)

maiages<-sample(c(45:54),550,replace=TRUE)

majages<-sample(c(55:64),237,replace=TRUE)
```

```
makages<-sample(c(65:74),94,replace=TRUE)

malages<-sample(c(75:84),70,replace=TRUE)

mamages<-sample(c(85:100),58,replace=TRUE)

massaultages<-c(maages,mabages,macages,madages,maeages,mafages,magages,mahages,maiages,majages,makages,malages,makages)


#look at 5 number summary and plots for assaults in Maricopa County

summary(massaultages)


#look at differences between all deaths and death by assault

boxplot(massaultages,mallages,

        main=c("Deaths in Maricopa County 1999-2015"),

        ylab=c("Age"),

        names=c("Assault","All Deaths"))


hist(mallages,

        main=c("All deaths in Maricopa County 1999-2015"),

        xlab=c("Age"),)


hist(massaultages,

        main=c("Assault deaths in Maricopa County 1999-2015"),

        xlab=c("Age"))


#Plot Q - Q plot of deaths

#This is not normal distribution

qqplot(sort(mallages),sort(massaultages),

        xlim=c(0,100),

        ylim=c(0,100),

        col="black",

        xlab = c("All Deaths"),

        ylab = c("Assault Deaths"),

        main=c("Q - Q plot of All vs Assault Deaths in Maricopa"))
```

```
abline(a=0,b=1)

q<-c(0.25,0.50,0.75)

points(quantile(mallages,q),quantile(massaultages,q),col=rainbow(3),pch=19)

legend("topleft",

        c("1st quantile","2nd quantile", "3rd quantile"),

        fill = c(rainbow(3)),

        cex=0.8)


#The two counties look similar and have simliar means.  Let's

#look at the deaths by assault between the two counties

pima=assault/pimadeath*100

mar=marassault/mardeath*100


qqplot(sort(pima),sort(mar),

        xlim=c(0,30),

        ylim=c(0,30),

        col="black",

        xlab = c("Pima Assault Deaths"),

        ylab = c("Maricopa Assault Deaths"),

        main=c("Q - Q plot of Assault Deaths in Pima vs Maricopa"))

abline(a=0,b=1)

q<-c(0.25,0.50,0.75)

points(quantile(pima,q),quantile(mar,q),col=rainbow(3),pch=19)

legend("bottomright",

        c("1st quantile","2nd quantile", "3rd quantile"),

        fill = c(rainbow(3)),

        cex=0.8)


#Run a t-test to see if the difference of the means of the percentage of

#people in each age group who die due to assault is statistically significant

#in Pima and Maricopa county Arizona.
```

#H0: There is no difference in the means of the percent of people who die from assault (as opposed to all causes of death) in Pima county and Maricopa county.

#H1: The percent of people who die from assault is greater in Maricopa county than in Pima county.

t.test(mar,pima,alternative=c("greater"),paired=TRUE)

RStudio

2022.07.1 Build 554

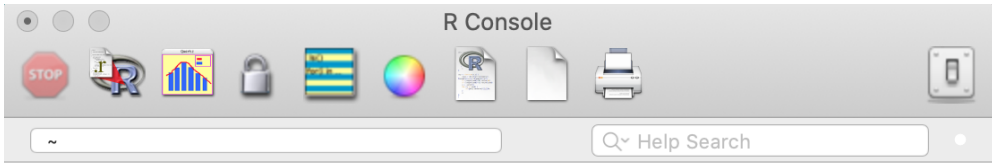
"Spotted Wakerobin" Release (7872775e, 2022-07-22) for macOS

Mozilla/5.0 (Macintosh; Intel Mac OS X 11_6_2) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.10 Chrome/69.0.3497.128 Safari/537.36

R Version:

RStudio 2022.07.1+554 "Spotted Wakerobin" Release (7872775ebddc40635780ca1ed238934c3345c5de, 2022-07-22) for macOS

Mozilla/5.0 (Macintosh; Intel Mac OS X 11_6_2) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.10 Chrome/69.0.3497.128 Safari/537.36




```
R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale
```