BIOST 515/518 Homework 4

Latera Tesfaye Olana

02 March, 2023

Responses:

Question 1:

Summary of the data is provided in table 1.

Table 1: Summary of our data

	Total
	(N=2709)
county	
Adams County, WA	53 (2.0%)
Asotin County, WA	77 (2.8%)
Benton County, WA	80 (3.0%)
Chelan County, WA	80 (3.0%)
Clallam County, WA	80 (3.0%)
Clark County, WA	80 (3.0%)
Columbia County, WA	13 (0.5%)
Cowlitz County, WA	80 (3.0%)
Douglas County, WA	78 (2.9%)
Ferry County, WA	26 (1.0%)
Franklin County, WA	80 (3.0%)
Garfield County, WA	3 (0.1%)
Grant County, WA	80 (3.0%)
Grays Harbor County, WA	80 (3.0%)
Island County, WA	80 (3.0%)
Jefferson County, WA	80 (3.0%)
King County, WA	80 (3.0%)
Kitsap County, WA	80 (3.0%)
Kittitas County, WA	79 (2.9%)
Klickitat County, WA	76 (2.8%)
Lewis County, WA	80 (3.0%)
Lincoln County, WA	53 (2.0%)
Mason County, WA	80 (3.0%)
Okanogan County, WA	80 (3.0%)
Pacific County, WA	80 (3.0%)
Pend Oreille County, WA	66 (2.4%)
Pierce County, WA	80 (3.0%)
San Juan County, WA	62 (2.3%)
Skagit County, WA	80 (3.0%)

	Total
Skamania County, WA	36 (1.3%)
Snohomish County, WA	80 (3.0%)
Spokane County, WA	80 (3.0%)
Stevens County, WA	80 (3.0%)
Thurston County, WA	80 (3.0%)
Wahkiakum County, WA	12 (0.4%)
Walla Walla County, WA	80 (3.0%)
Whatcom County, WA	80 (3.0%)
Whitman County, WA	75 (2.8%)
Yakima County, WA	80 (3.0%)
gender	
Female	1325 (48.9%)
Male	1384 (51.1%)
age	
55-64 years	617 (22.8%)
65-74 years	683 (25.2%)
75-84 years	695 (25.7%)
85+ years	714 (26.4%)
deaths	
Mean (SD)	172 (298)
Median [Min, Max]	69.0 [10.0, 3070]
pop	
Mean (SD)	7050 (14900)
Median [Min, Max]	2510 [35.0, 134000]

Table 1 provides summary of the datasets. In the table I have only selected to show variables, which can be expressed in-terms mean, median and count (proportion). All variables expect year are summarized in the table. From this table we can tell which counties has the most and the least death report recorded. In total we expect (4 - age * 2 - gender * 10 - years), 80 data points. We can show the years, gender group or age group certain counties did had no deaths reported. This becomes important if we are interested in knowing if the missing of the report is related to no death being observed or not reporting deaths with despite deaths being observed. For instance, Wahkiakum County, WA county only reported 12 data points (deaths for each gender and age groups and year).

Table 2: The first five observations ordered by population

county	gender	age	year	deaths	pop	emprical_rate
Garfield County, WA	Male	85+ years	2018	10	35	28571.43
Ferry County, WA	Male	85+ years	2016	12	47	25531.91
Ferry County, WA	Male	85+ years	2010	15	49	30612.24
Ferry County, WA	Male	85+ years	2011	13	49	26530.61
Wahkiakum County, WA	Male	85+ years	2015	10	52	19230.77

Table 2 shows the first 5 observations ordered by population variable to check if there are any entry error. In addition, *is.na()*, *min()*, *max()*, *as.factor()*, *arrange()* and *typeof()* are used to assess if there were any errors in the dataset. There are no missing values. One peculiar pattern I have noticed is when the data is order by deaths, we get in 2018 from 664 Adams County, WA, Female 65-74 years of age population 10 deaths. However, changing the variable on which we arrange the data to population varaiable completely provides different picture. In 2018 from 35 Garfield County, WA Male 85+ years age of population 10 deaths were recorded. This implies rather huge difference in mortality rate. I am not sure to weather this is a data

entry error or related some other factors, therefore in subsequent analysis I will use the dataset as provided. Empirical mortality per 100,000 thousand population was estimated by dividing the number of deaths by the number of population.

I will start the explanatory analysis from obvious visualization. I will plot the crude deaths on the y-axis and population number on the x-axis. The idea being for close mortality rate between groups or counties, the proportion between population and death should much. Accordingly, figure 1 shows, for age group above 85+ we can see high number of deaths for proportional similar population number as compared to other age groups. The color in the figures represents states and it can be seen state(s) represented by green color have high number of deaths recorded across all four age groups.

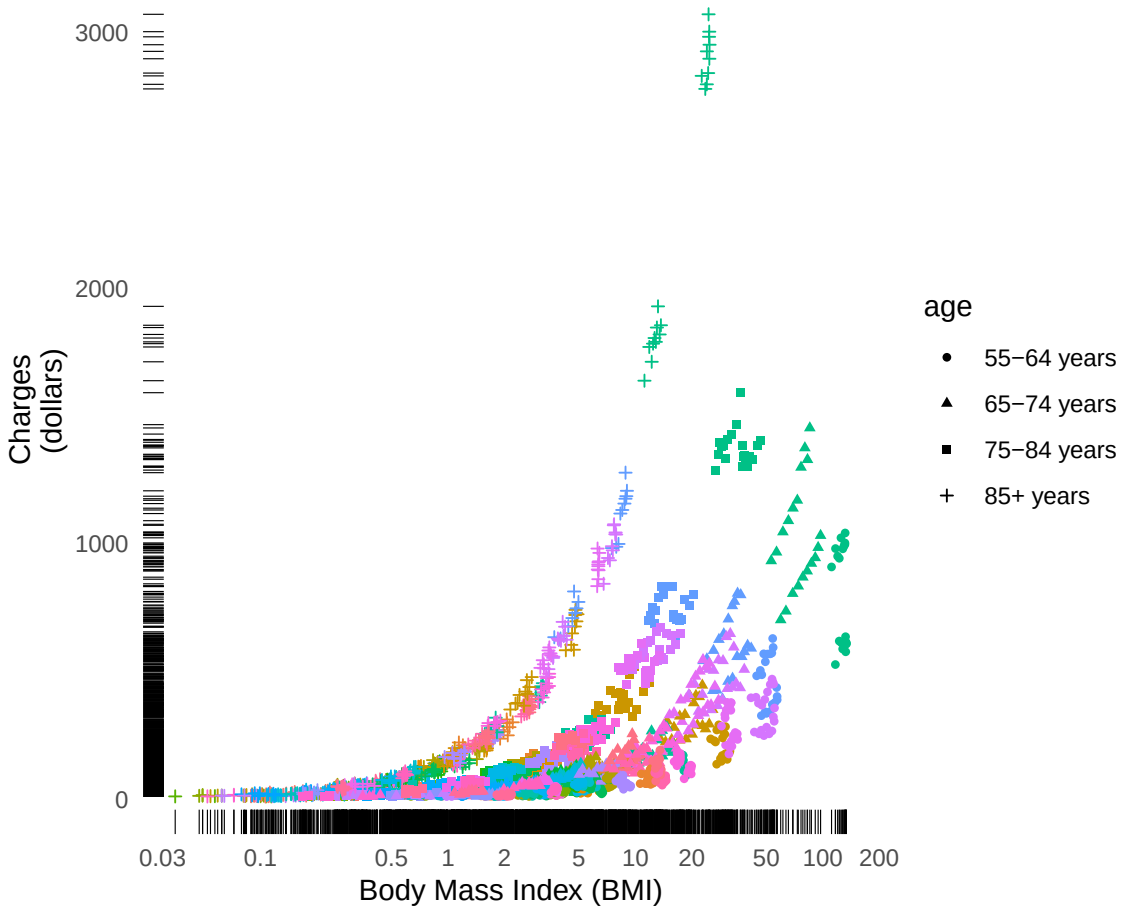


Figure 1: Plotting deaths with population (imbalance in death can be seen from this plot)

Figure 2 shows the relationship between age, gender and empirical mortality rate per 100,000 population across time (from 2010 to 2019). Generally, across the state male have highest mortality rate. For age group 55 - 64 years of age for both genders we can see an increasing trend. For 65 - 74 age groups the mortality rate somehow stayed constant for males, however decreased for female. For age group 75 - 84 years of age for both genders the mortality rate decreased. Lastly, for 85+ age group we can see large variability in number of deaths and it showed slight decrease in male, whereas, in female we can see slight increase in the mortality rate. This complementary to the above graphs as there are some states with significantly larger death in 85+ years of age female population.

We can look at this data from another visualization perspective as well. For instance figure 3, shows on the x-axis the given age group and on the y-axis empirical mortality rate per 100,000 population. There is an

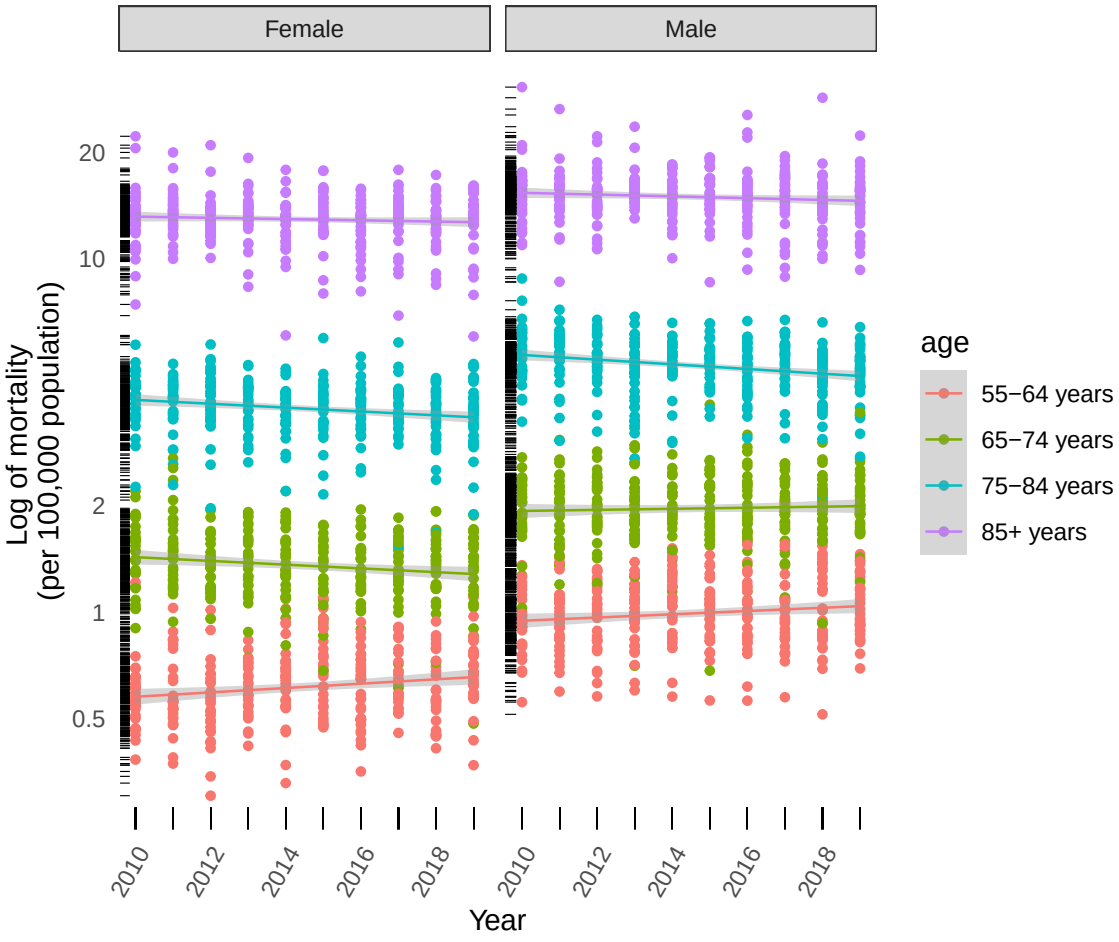


Figure 2: Age, gender and year plot for showing mortality rate

increasing trend in mortality rate across age groups (from 54 - 65 to 85+).

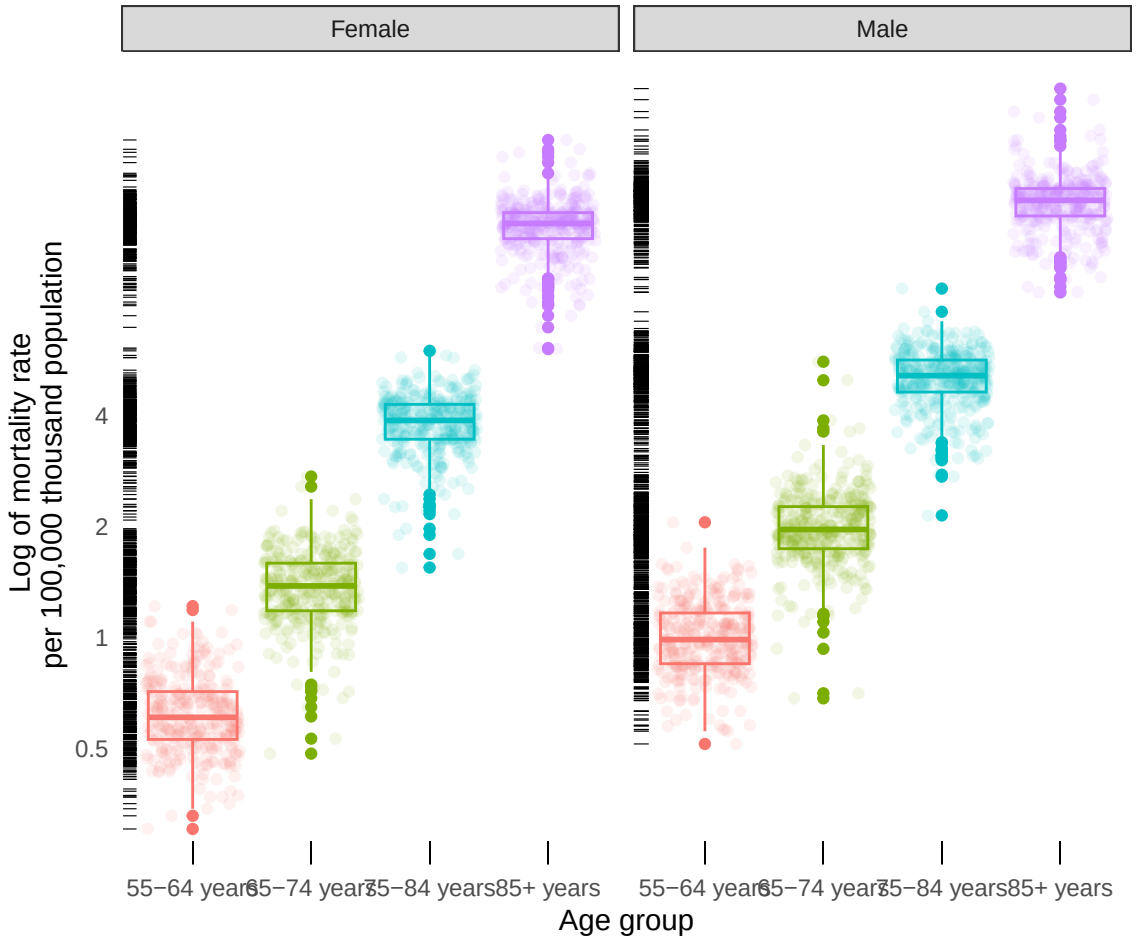


Figure 3: Mortality rate (per 100000) across age groups

Finally, we can borrow figure 1 to assess generally if there is increasing or decreasing trend in mortality rate across years. As shown in figure 4, there seems to be a slight decrease/increase but it is very small. The second question I had was is the above relationships between mortality rate and the given covarities also the same if we drill down to county level. In the annex two supplementary plots were provided.

Question 2

The appropriate parameter variable for ti (exposure window) selection in this case would be the population size (pop) for each county, sex, age, and year. Since the mortality rate is defined as the number of deaths per unit of population, the population size is the appropriate variable to use as the denominator in the mortality rate calculation. Using population size as the exposure window variable ensures that the mortality rates are calculated based on a consistent population size across different counties, sexes, age groups, and years, and allows for valid comparisons of mortality rates over time and across demographic groups.

Let's start from our simple model.

$$EY_i = t_i * \lambda_i$$

In the above formula Y_i The annual number of deaths observed in counties of Washington (WA). t_i is exposure window (here we will be using the total population of certain age group and gender in each counties of WA). λ_i is the annual mortality rate. Rate here is defined as number of observed deaths divided by the number of population for a given age group and gender.

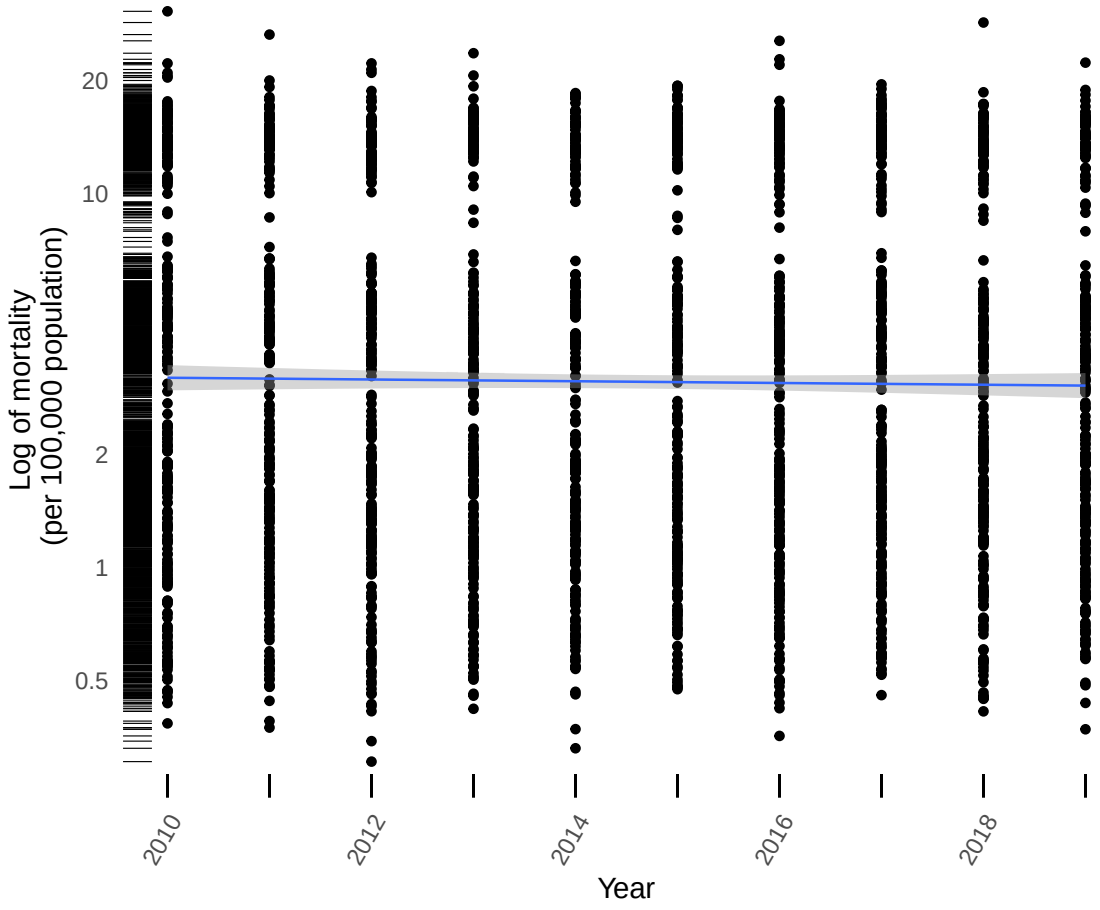


Figure 4: Over all plot of change in mortality across year

Know we can define $log(\lambda_i)$, which is given by using the given covariates. Accordingly, $log(\lambda_i) = \beta_0 + \beta_1 \times 1_{gender_i:if'i'isformale} + \beta_3 \times 1_{age_i:for'i'is65-74} + \beta_4 \times 1_{age_i:for'i'is75-84} + \beta_5 \times 1_{age_i:for'i'is85_{above}} + \beta_6 \times year$

Putting this all together the model we want to fit is given below:

$log(E[deaths_i|county_i, gender_i, age_i, year_i]) = \beta_0 + \beta_1 \times 1_{gender_i:if'i'isformale} + \beta_2 \times 1_{age_i:for'i'is65-74} + \beta_3 \times 1_{age_i:for'i'is75-84} + \beta_4 \times 1_{age_i:for'i'is85_{above}} + \beta_5 \times year$

At our baseline, we have age group from 54 - 65, female gender, and year 2010. To facilitate interpretation, I will scale the year variable so that our baseline is 2010.

The fitted model can be re-written as follow:

$log(E[\hat{deaths}_i|county_i, gender_i, age_i, year_i]) = log(pop_i) + \hat{\beta}_0 + \hat{\beta}_1 \times 1_{gender_i:if'i^{th}observation'isformale} + \hat{\beta}_2 \times 1_{age_i:for'i^{th}observation'is65-74} + \hat{\beta}_3 \times 1_{age_i:for'i^{th}observation'is75-84} + \hat{\beta}_4 \times 1_{age_i:for'i^{th}observation'is85_{above}} + \hat{\beta}_5 \times year$

Where, $log(E[\hat{deaths}$ is the estimated log mean mortality rate of the i^{th} observation group. $log(pop_i)$ is the offset value. (The description of the variables are included in the model)

Characteristic	log(IRR)	95% CI	p-value
(Intercept)	-5.02044	-5.02983, -5.01106	<0.001
Gender			
Female	—	—	
Male	0.28694	0.28115, 0.29274	<0.001
Age			
55-64 years	—	—	
65-74 years	0.76277	0.75295, 0.77260	<0.001
75-84 years	1.76968	1.76033, 1.77904	<0.001
85+ years	2.94561	2.93674, 2.95450	<0.001
Year	-0.00371	-0.00471, -0.00270	<0.001

Using Poisson regression, $\hat{\beta_0}$ = -5.02 is the log estimated death rate (per capita per year) for a group of individuals who their gender are provided as female and aged 54 to 65 years of age for the year 2010. Among this population, we estimate that the death rate is 660 cases per 100,000 people per year. (any reference to gender hereafter is as 'provided')

Using Poisson regression, $\hat{\beta_1} = 0.28694$ is the estimated difference in log mean death rate, comparing a male population to a female population (first minus second) adjusting for age and year. Alternatively, comparing these two populations, with one population female and the other male, we estimate that the male population has death rate that 1.33 times that of or 33% higher than the female population adjusting for age and year. (adjusting: for the same age and year)

Using Poisson regression, $\hat{\beta_2} = 0.76277$ is the estimated difference in log mean death rate, comparing a population aged 64 - 75 years of age to a population aged 54 - 65 years of age (first minus second) adjusting for gender and year. Alternatively, comparing these two populations, with one population aged 64 - 75 years of age and the other aged 54 - 65 years of age, we estimate that a population aged 64 - 75 years of age has death rate that is 2.14 times higher than the population aged 54 - 65 years of age adjusting for gender and year. (adjusting: for the same gender and year)

Using Poisson regression, $\hat{\beta_3}$ = -0.00371 is the estimated difference in log mean death rate, comparing two populations from two different years, which differ by one year adjusting for gender and age. Accordingly, we estimate the death rate is 0.996 times higher in a given age as compared to previous year holding gender and age constant. (adjusting: for the same gender and age)

Question 3

Using Poisson regression a 95% likelihood ratio-based confidence interval for the fold-difference in death rates between two years that differ in one year is 0.995-fold to 0.997-fold. Using non-robust hypothesis testing we reject the null hypotheses that death rates over the years are equal (likelihood ratio test $p < 0.001$).

Question 4

Characteristic	log(IRR)	95% CI	p-value
(Intercept)	-5.05984	-5.07463, -5.04510	<0.001
Gender			
Female	—	—	
Male	0.28698	0.28119, 0.29277	<0.001
Age			
55-64 years	—	—	
65-74 years	0.80580	0.78666, 0.82495	<0.001
75-84 years	1.85731	1.83947, 1.87518	<0.001
85+ years	2.96416	2.94723, 2.98112	<0.001
Year	0.00479	0.00216, 0.00742	<0.001
Age * Year			
65-74 years * Year	-0.00924	-0.01269, -0.00579	<0.001
75-84 years * Year	-0.01876	-0.02202, -0.01550	<0.001
85+ years * Year	-0.00398	-0.00707, -0.00088	0.012

We fit a Poisson regression model for log mortality rate, including age, gender, and year as indicator, and an interaction between year and age. We are assuming age might modify the relation between mortality rate and year.

Comparing two populations from two consecutive years, separated by one year, we estimate that the difference in log mean death rate comparing these two populations (latest minus earliest) is 0.99080 log units higher if both groups are aged between 65 - 74 years of age than if both groups are age between 55 - 64 years of age, for a population of the same gender. At the 5% level, we fail to reject the null hypothesis that there are no difference between these two groups (based on a 95% likelihood ratio-based confidence interval: (0.98739 to 0.99422); $p < 0.001$).

Comparing two populations from two consecutive years, separated by one year, we estimate that the difference in log mean death rate comparing these two populations (latest minus earliest) is 0.98141 log units higher if both groups are aged between 75-84 years of age than if both groups are age between 54 - 65 years of age, for a population of the same gender. At the 5% level, we fail to reject the null hypothesis that there are no difference between these two groups (based on a 95% likelihood ratio-based confidence interval: (0.97822 to 0.98462); $p < 0.001$).

Comparing two populations from two consecutive years, separated by one year, we estimate that the difference in log mean death rate comparing these two populations (latest minus earliest) is 0.99603 log units higher if both groups are aged between 85+ years of age than if both groups are age between 54 - 65 years of age, for a population of the same gender. At the 5% level, we fail to reject the null hypothesis that there are no difference between these two groups (based on a 95% likelihood ratio-based confidence interval: (0.99295 to 0.99912); $p < 0.012$).

Therefore, trends in mortality rates over time are typically different for different age groups.

Question 5

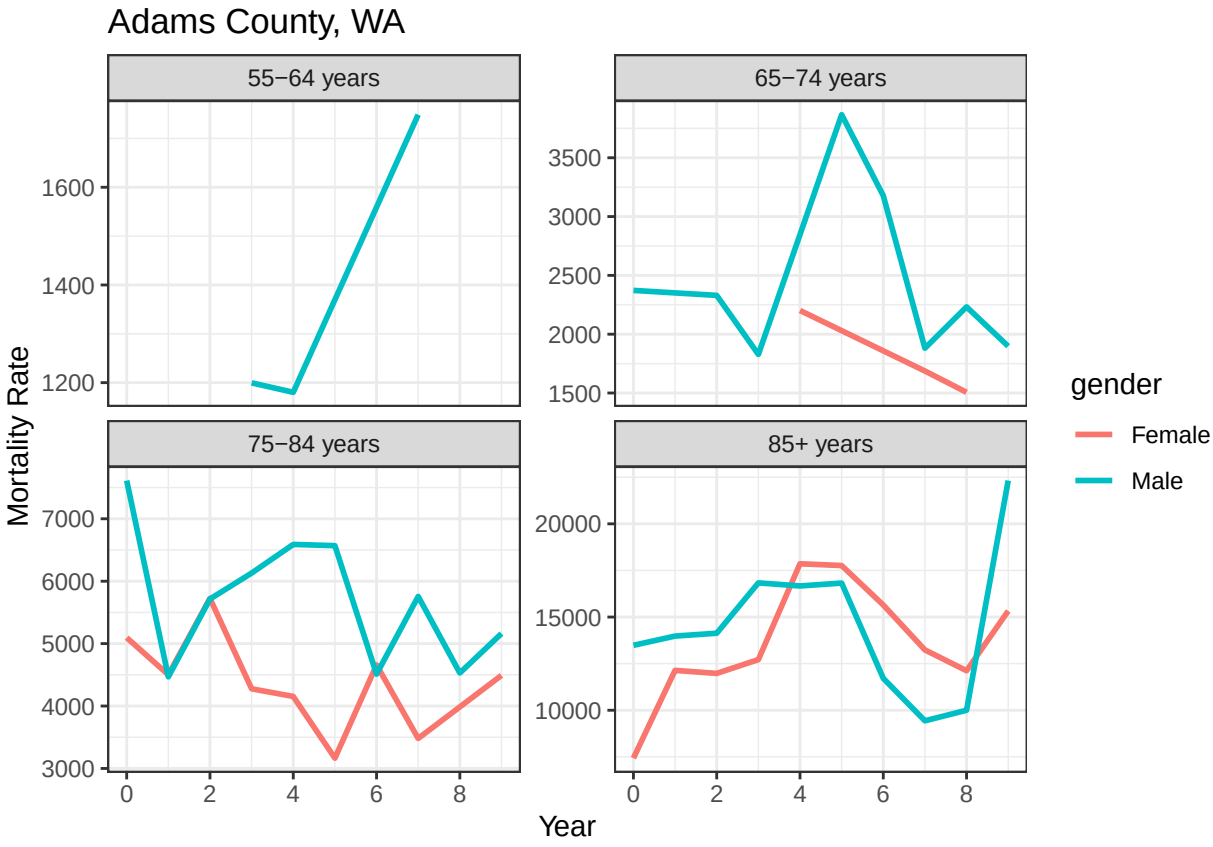
The assumption of independent observations is important in this type of analysis, as it enables reasonable estimates and valid statistical inference. However, in the above analyses, the assumption of independence may not be entirely reasonable due to the possibility of a cohort effect. As indicated in the description, the data covers the years 2010 to 2019, which means that some individuals may have been in different age groups during their lifetime. Broadly speaking, the assumption of independence may not be entirely reasonable due to the overall correlation between mortality rates within the same county, gender, age group, and year. For instance, counties with similar demographic and environmental characteristics may have similar mortality rates.

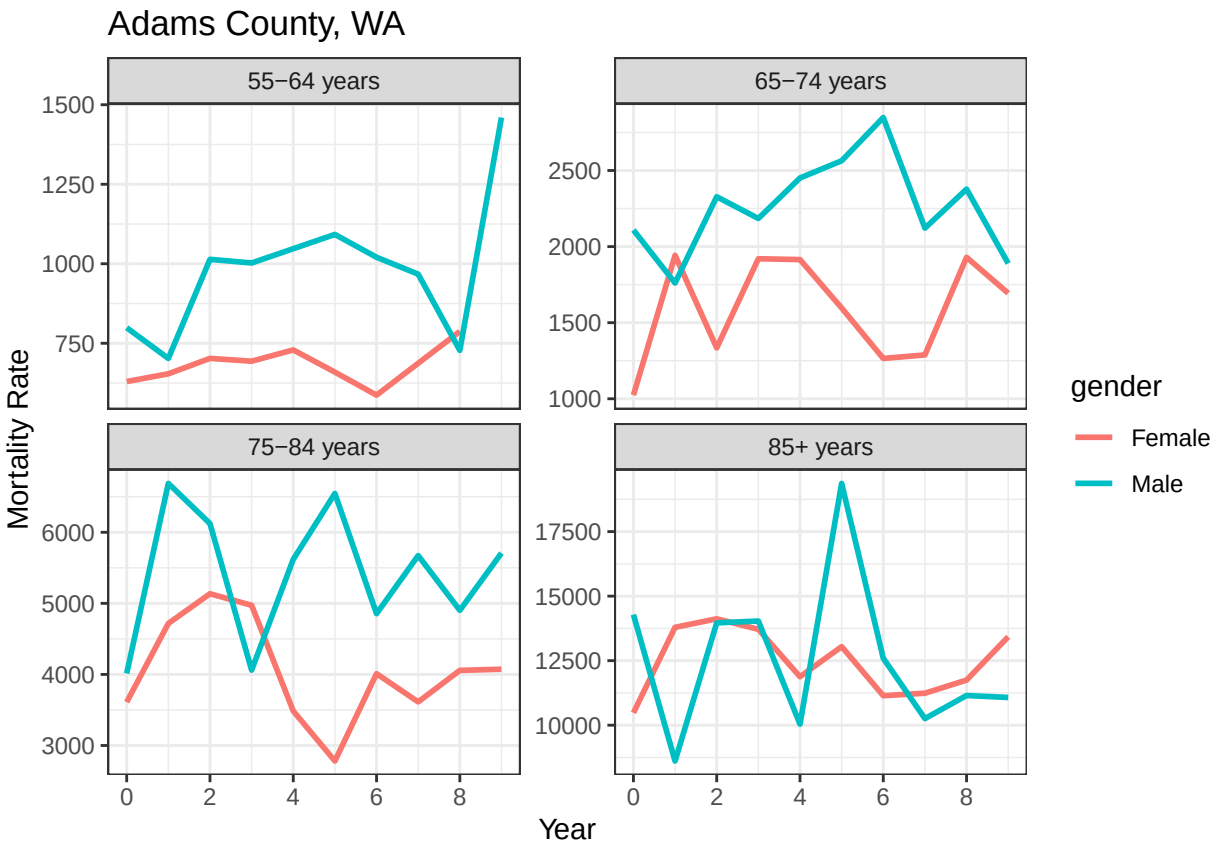
To mitigate the potential violation of independence, we can use advanced modeling schemes.

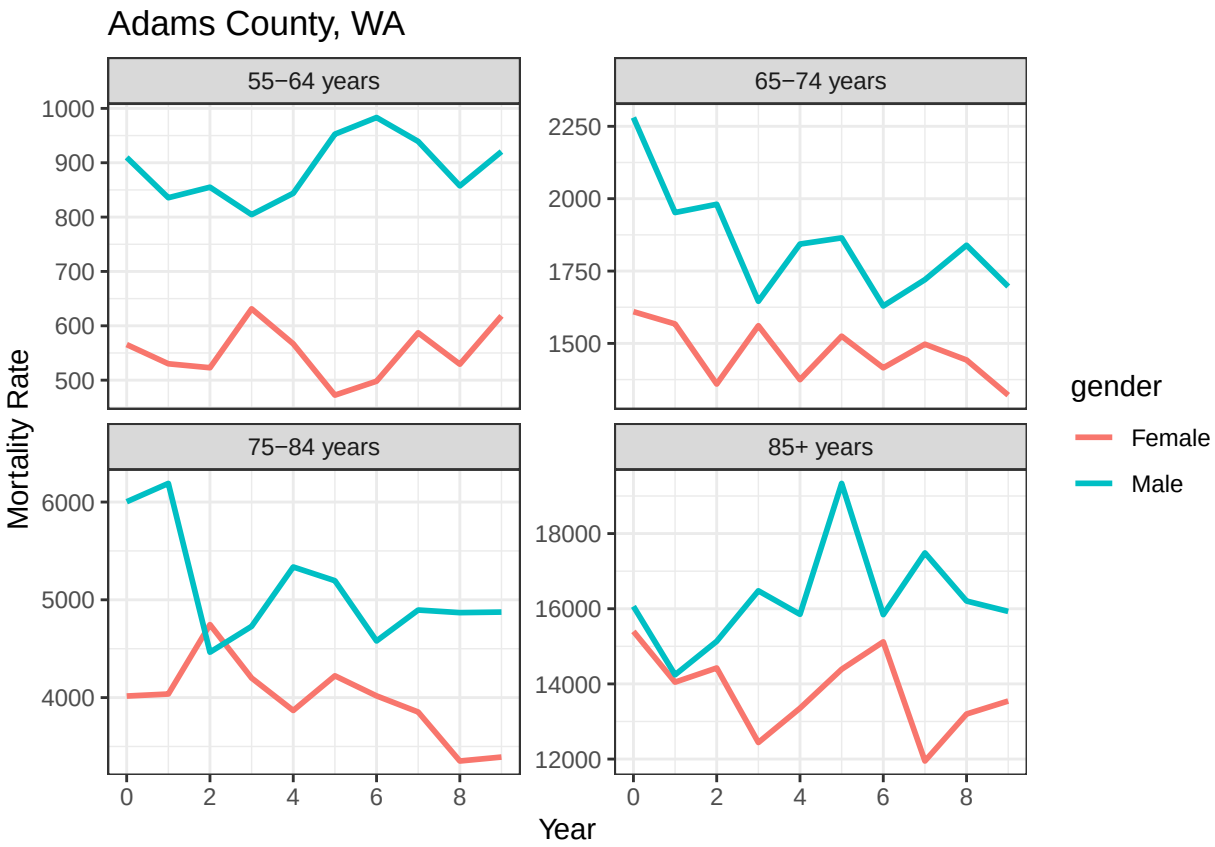
Annex - additional plots

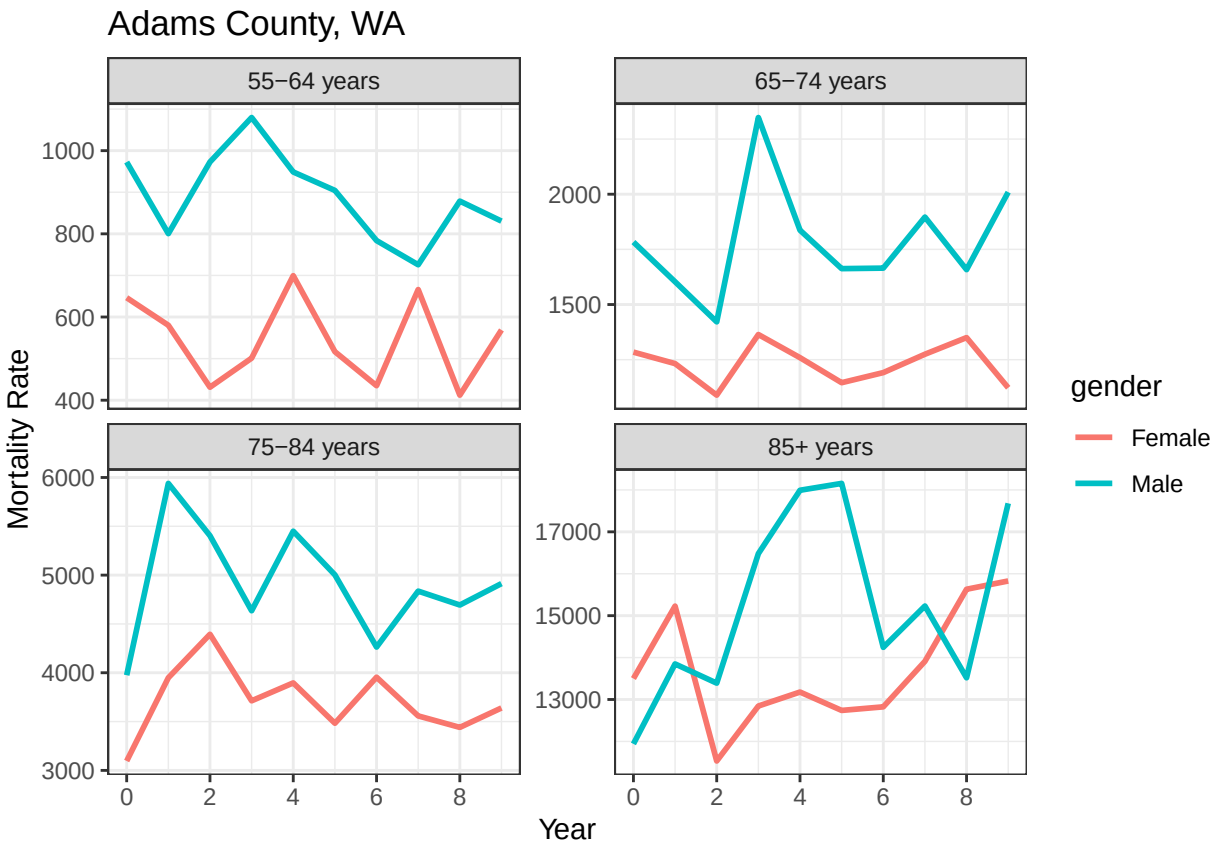
Plotting a time series of mortality rate per 100,000 population for all the county

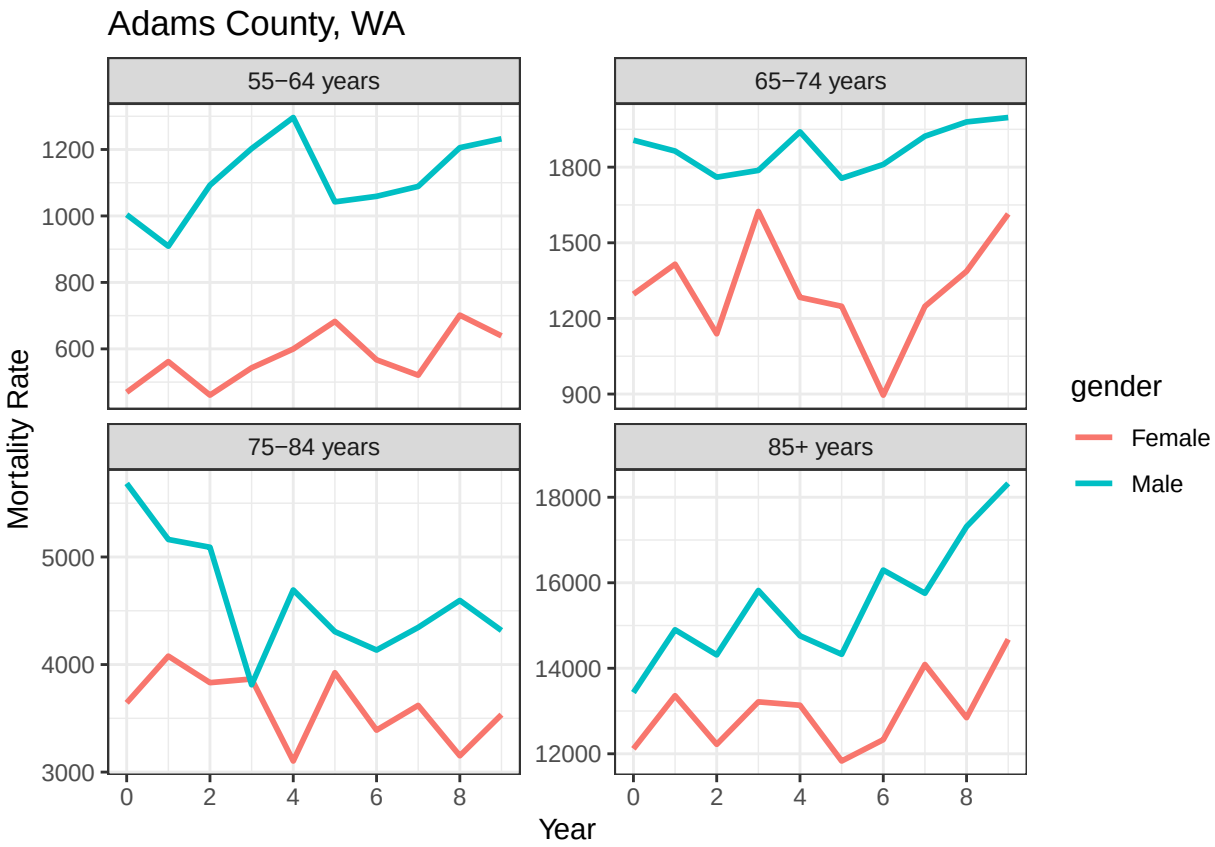
There is not much information to extract from this graph apart from seeing how mortality rate varied across years for each county.

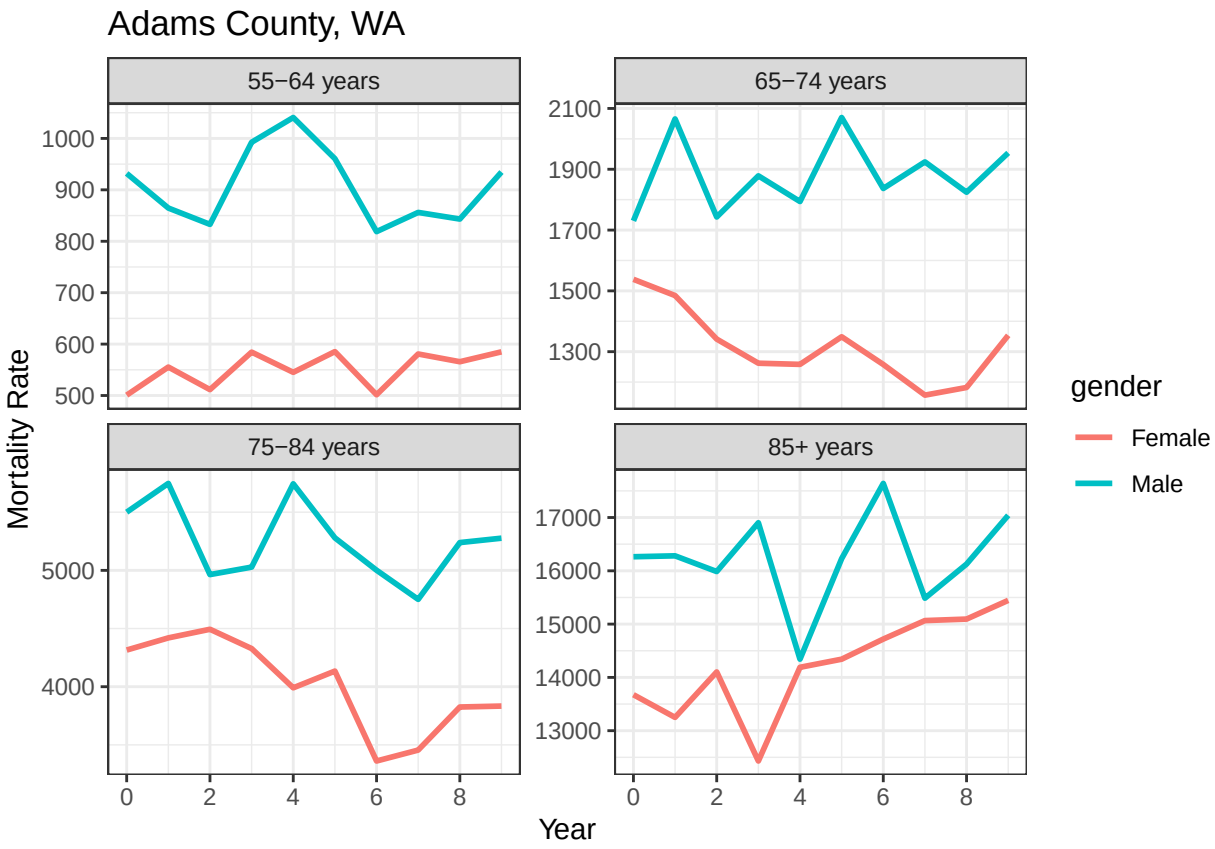


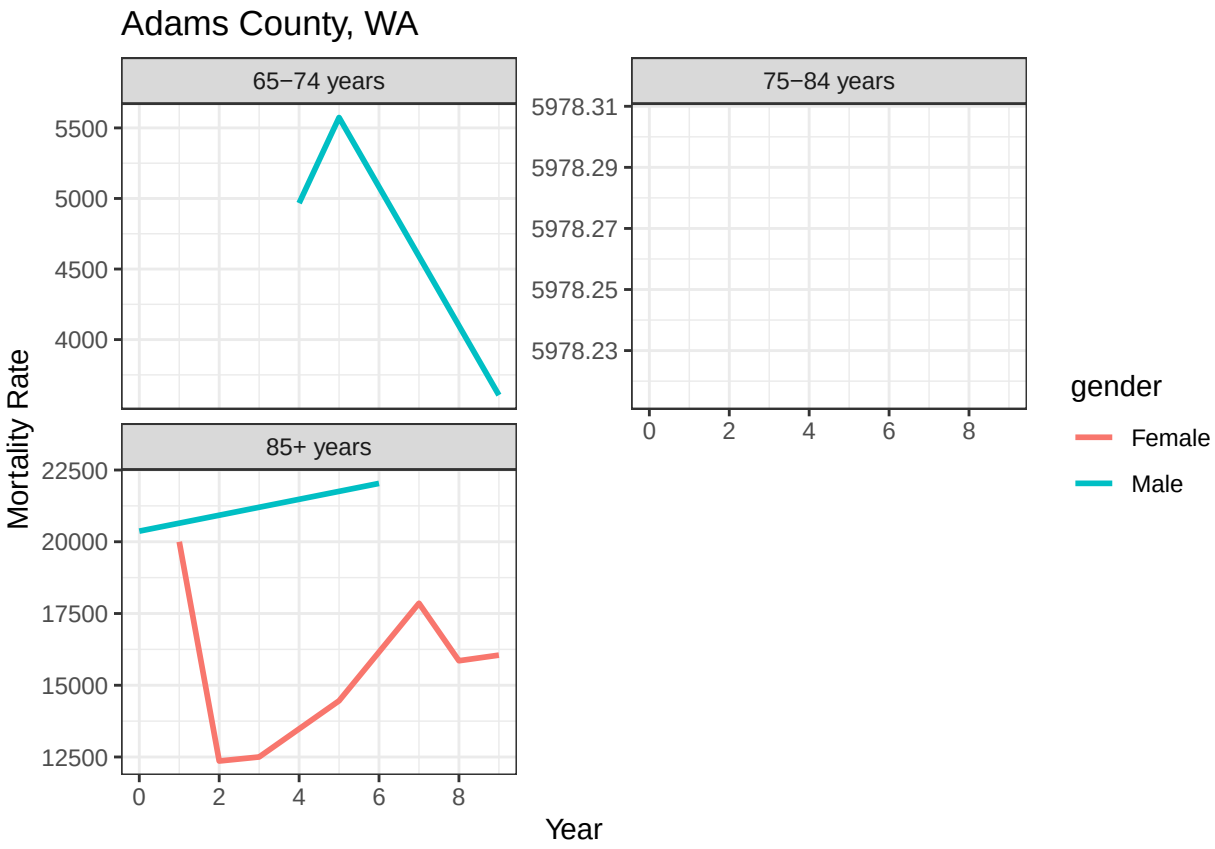

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

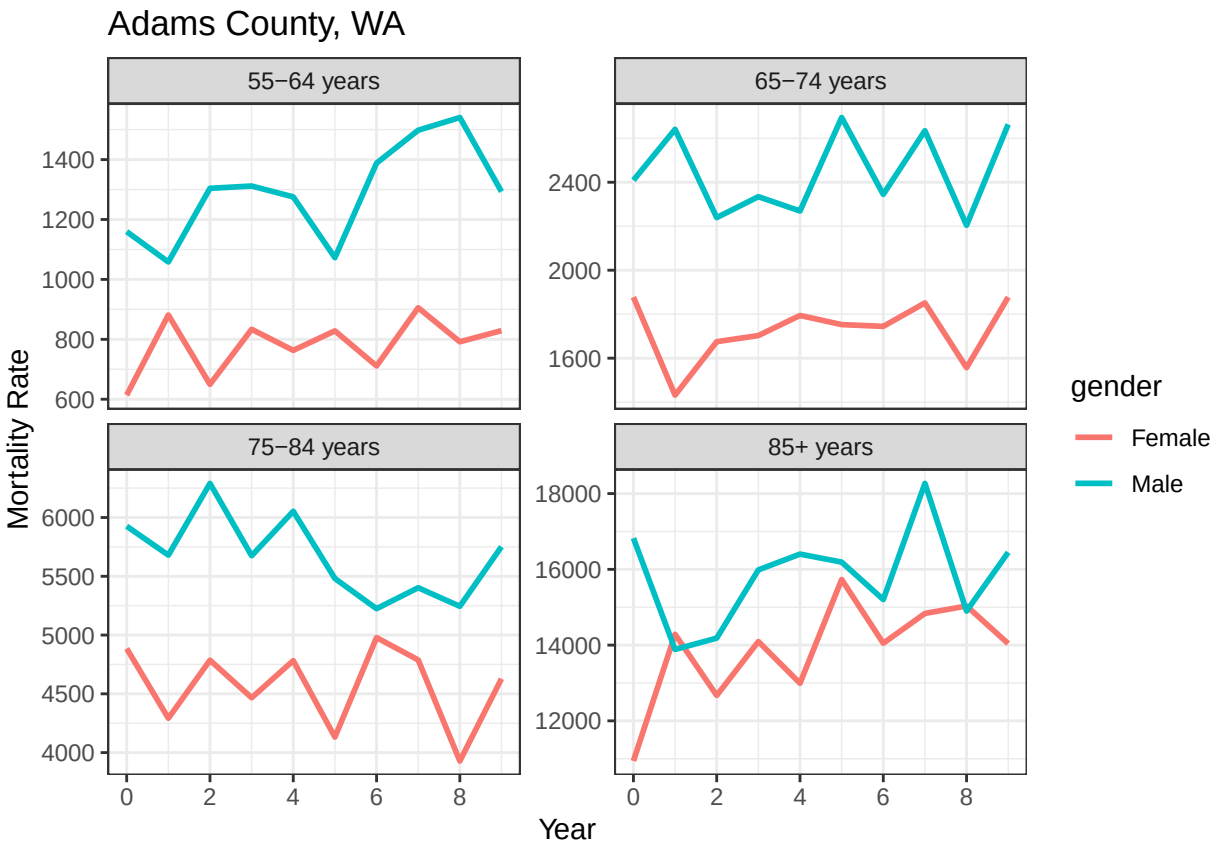

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

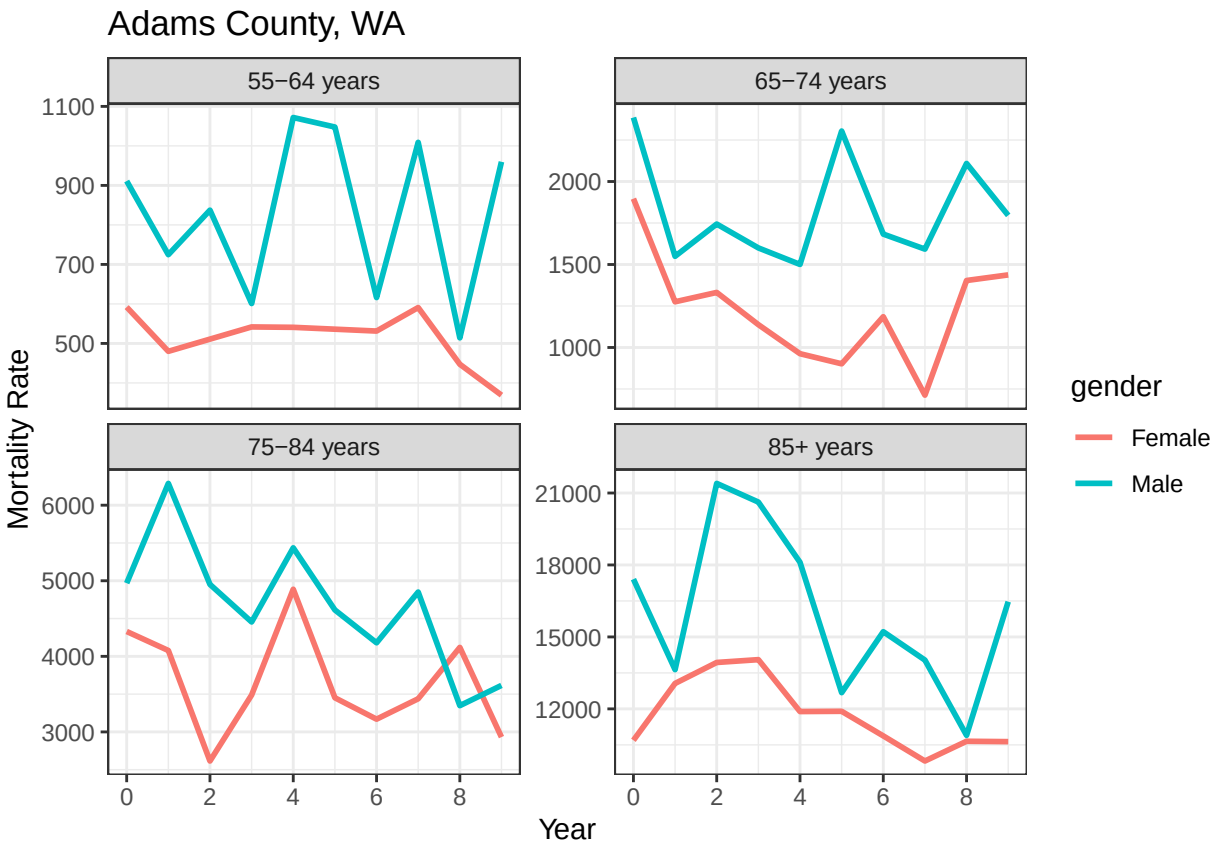

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

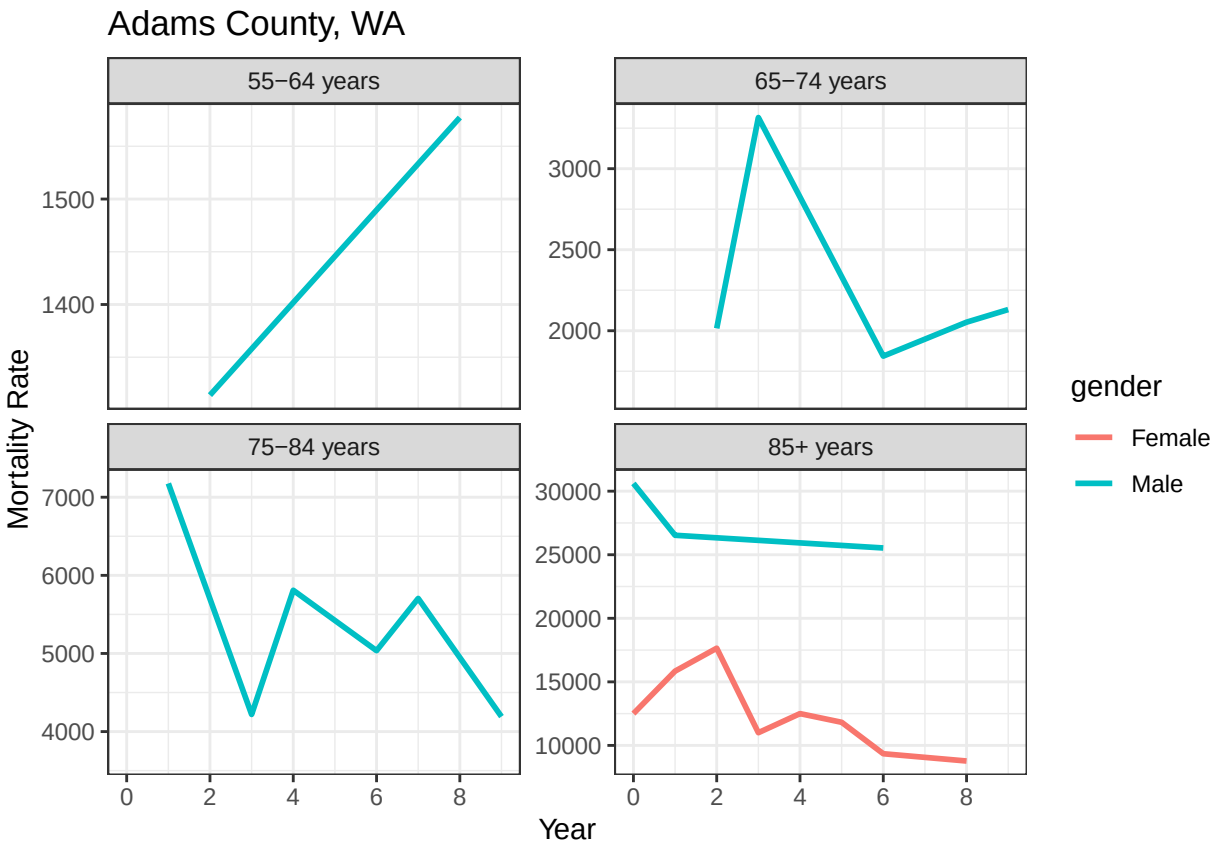

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

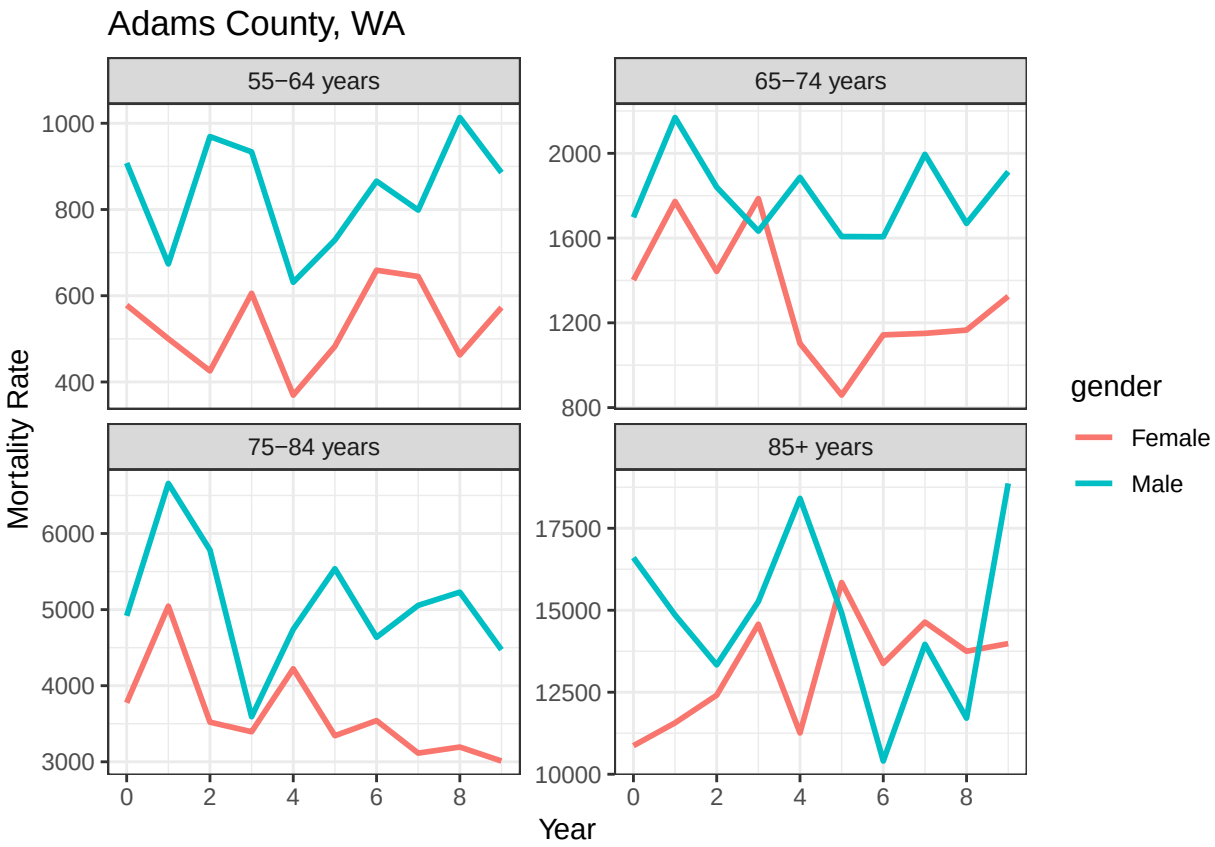

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

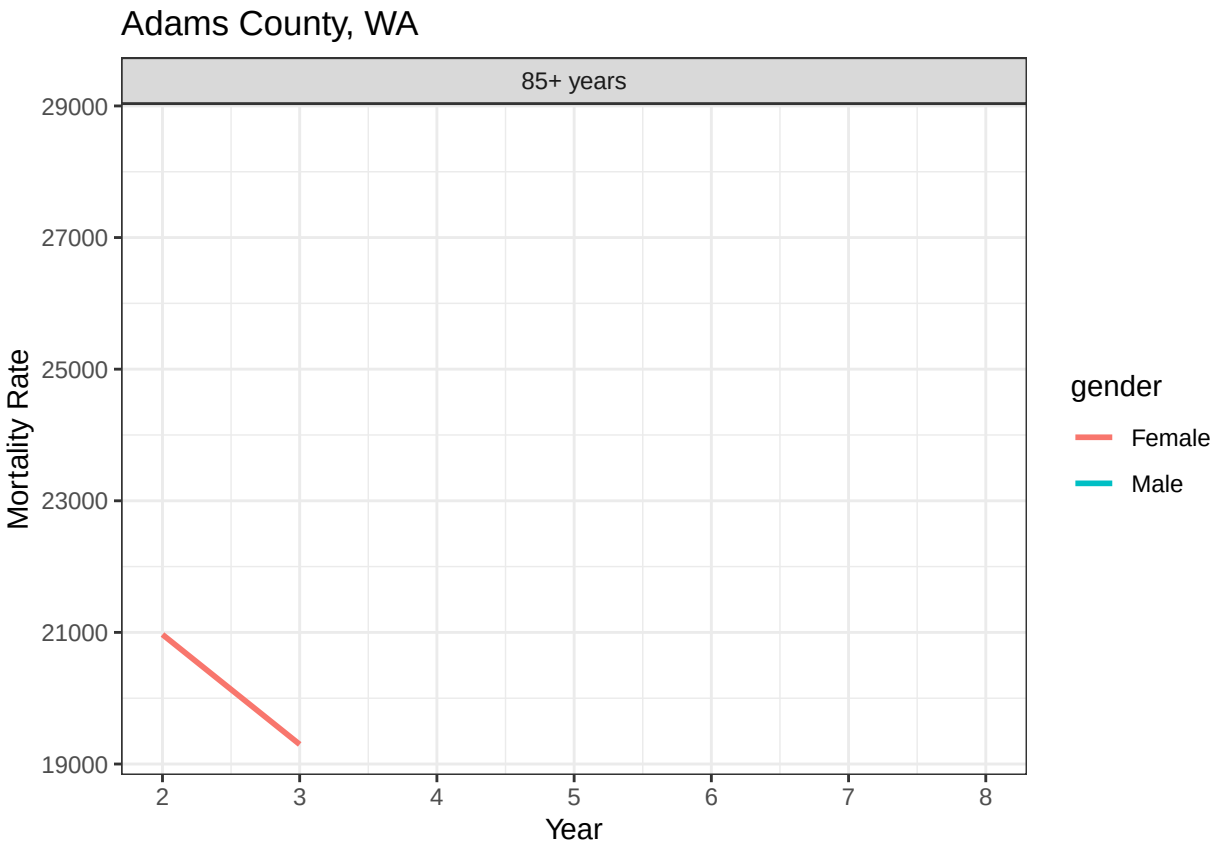

Adams County, WA
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

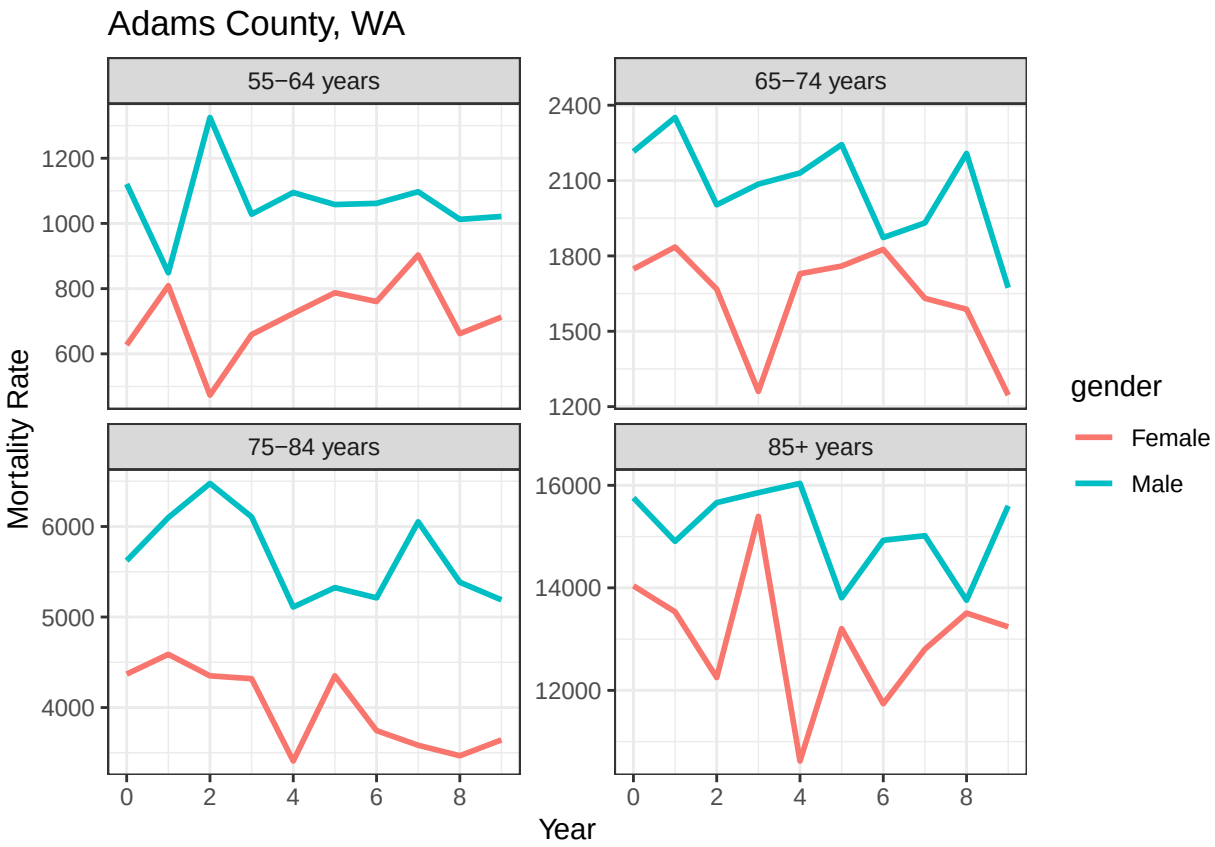

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

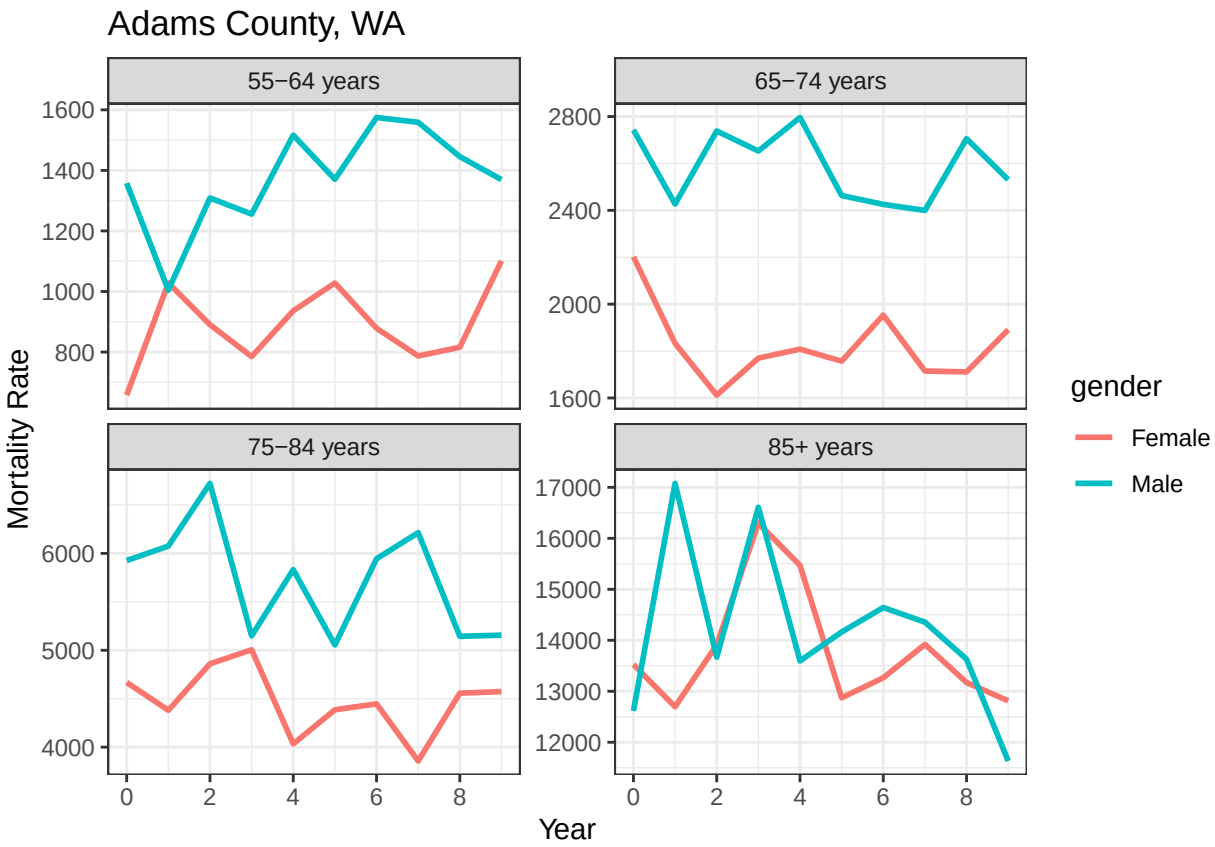

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

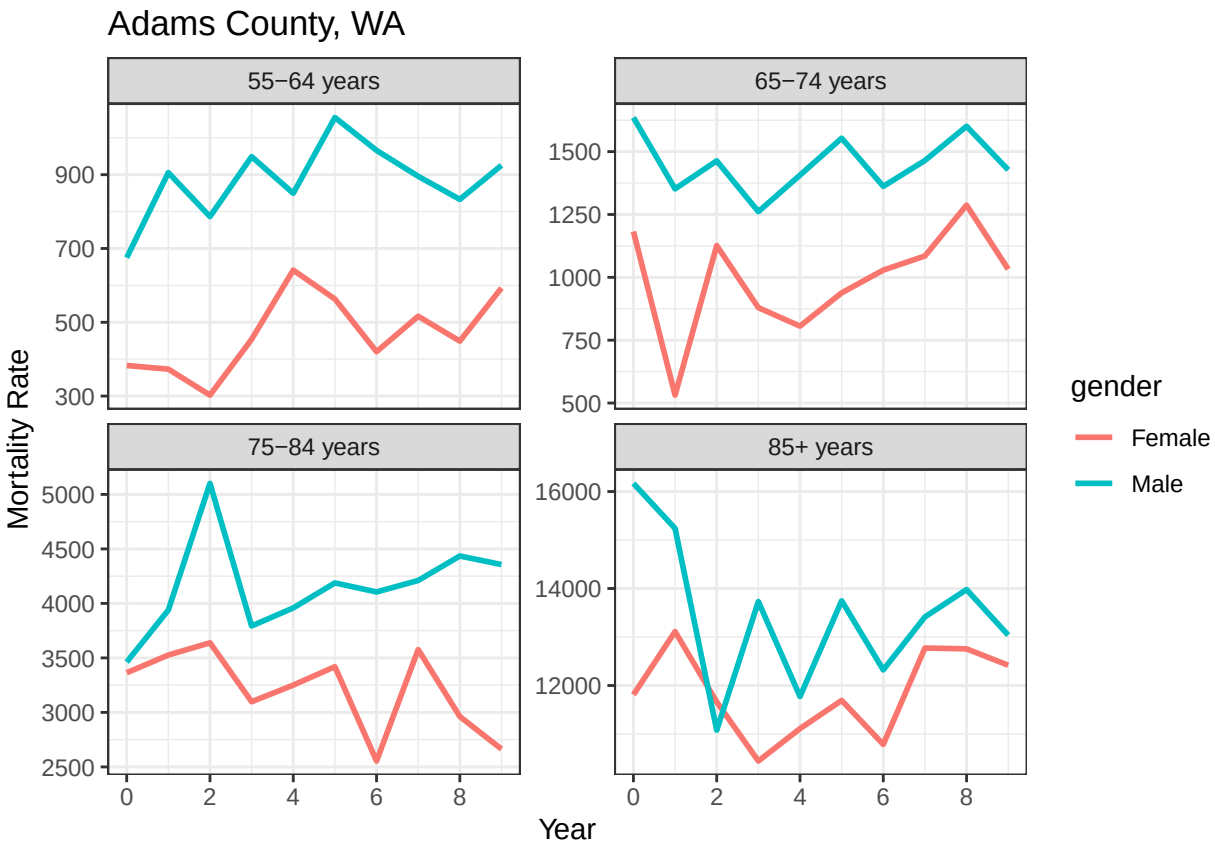

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male
1500
1400
3000
2500
2000
7000
6000
5000
4000
30000
25000
20000
15000
10000
0
2
4
6
8

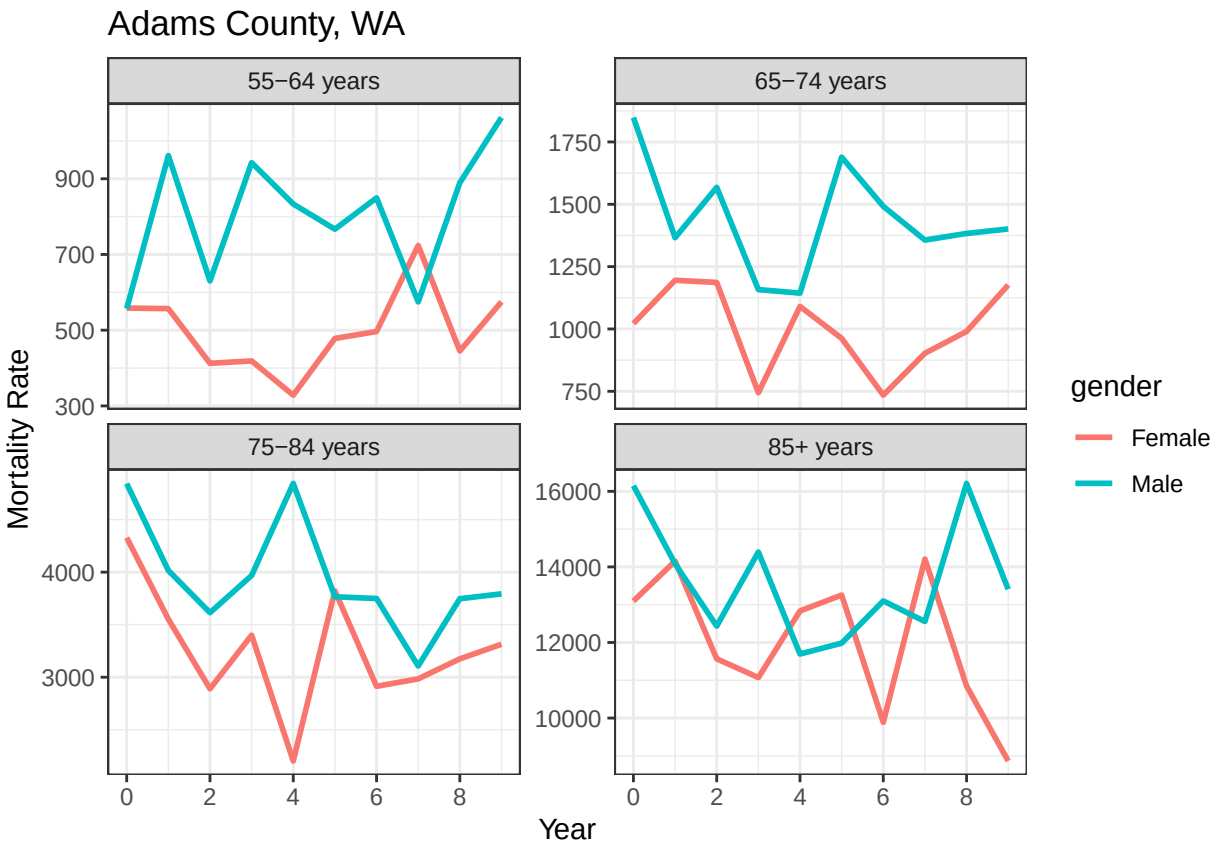

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

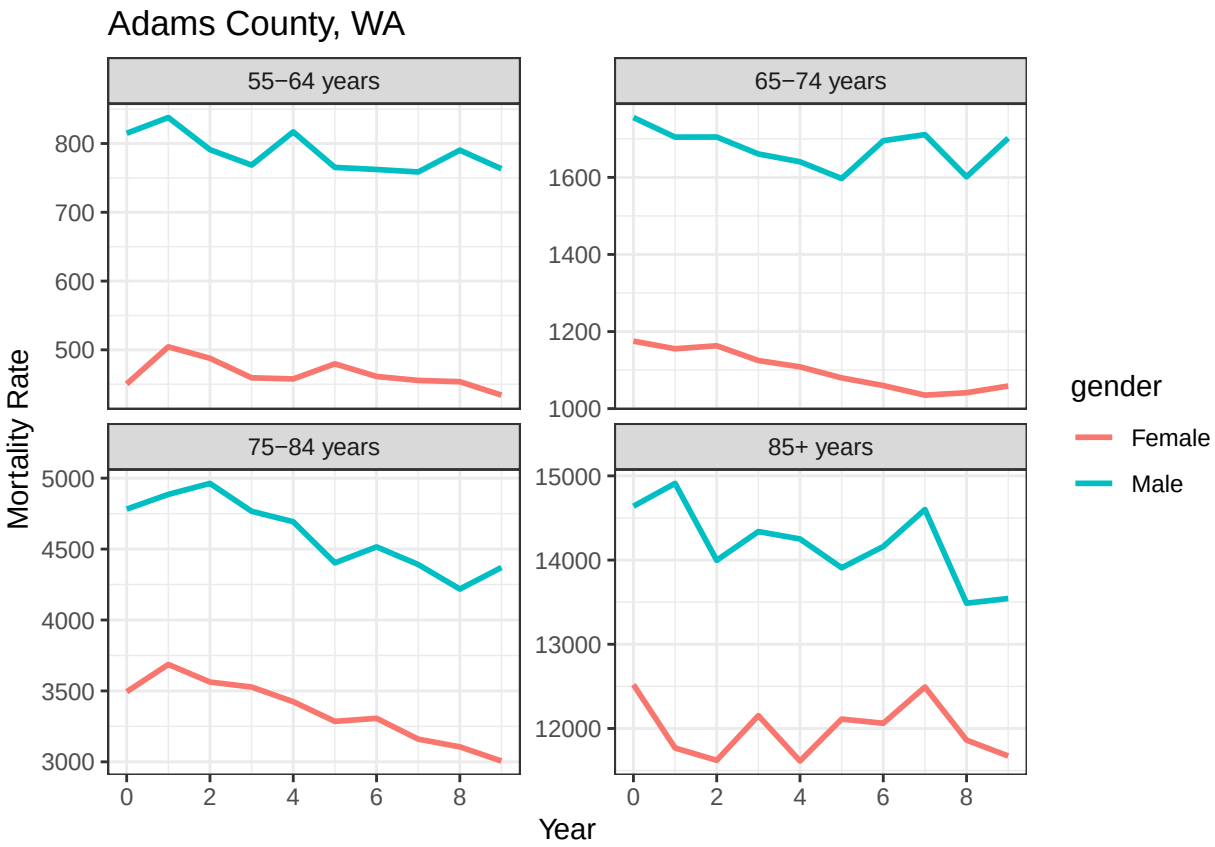

Adams County, WA
85+ years
29000
27000
25000
23000
21000
19000
Mortality Rate
2
3
4
5
6
7
8
Year
gender
Female
Male

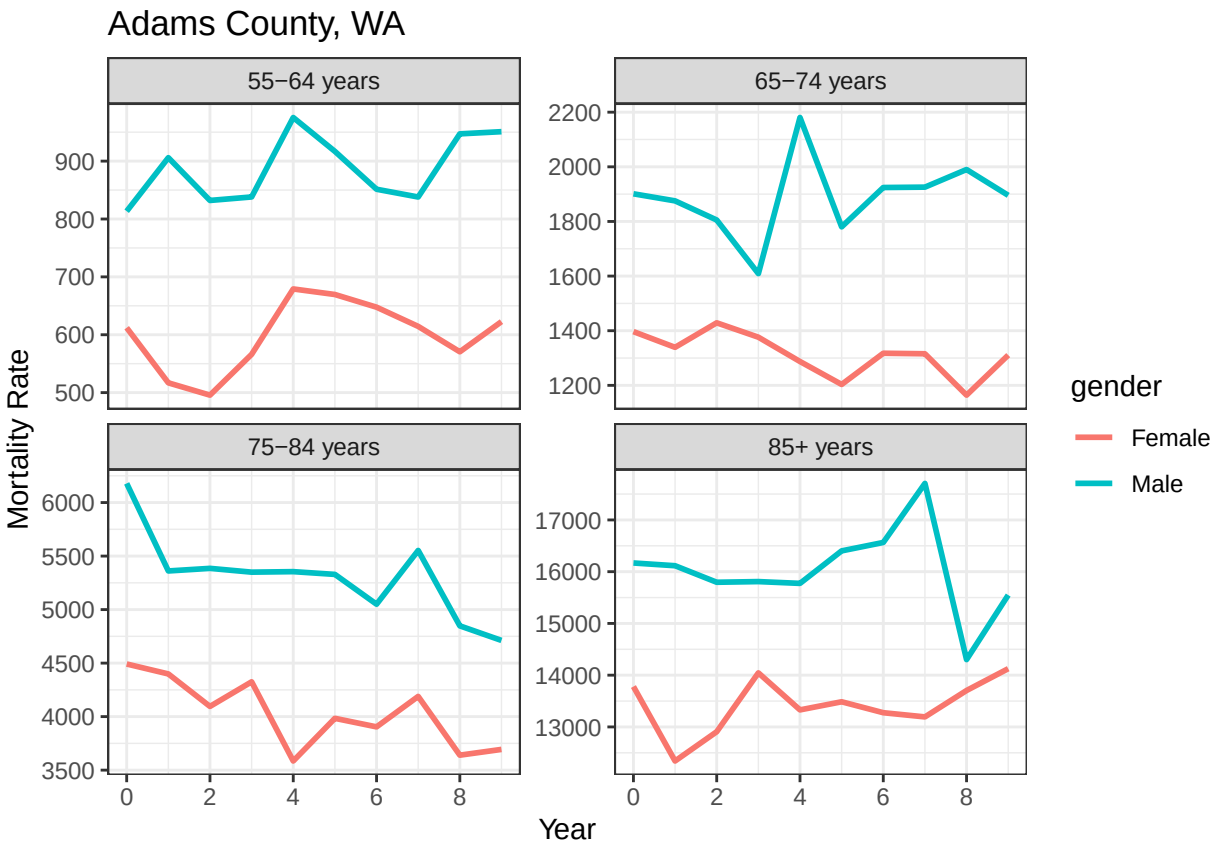

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

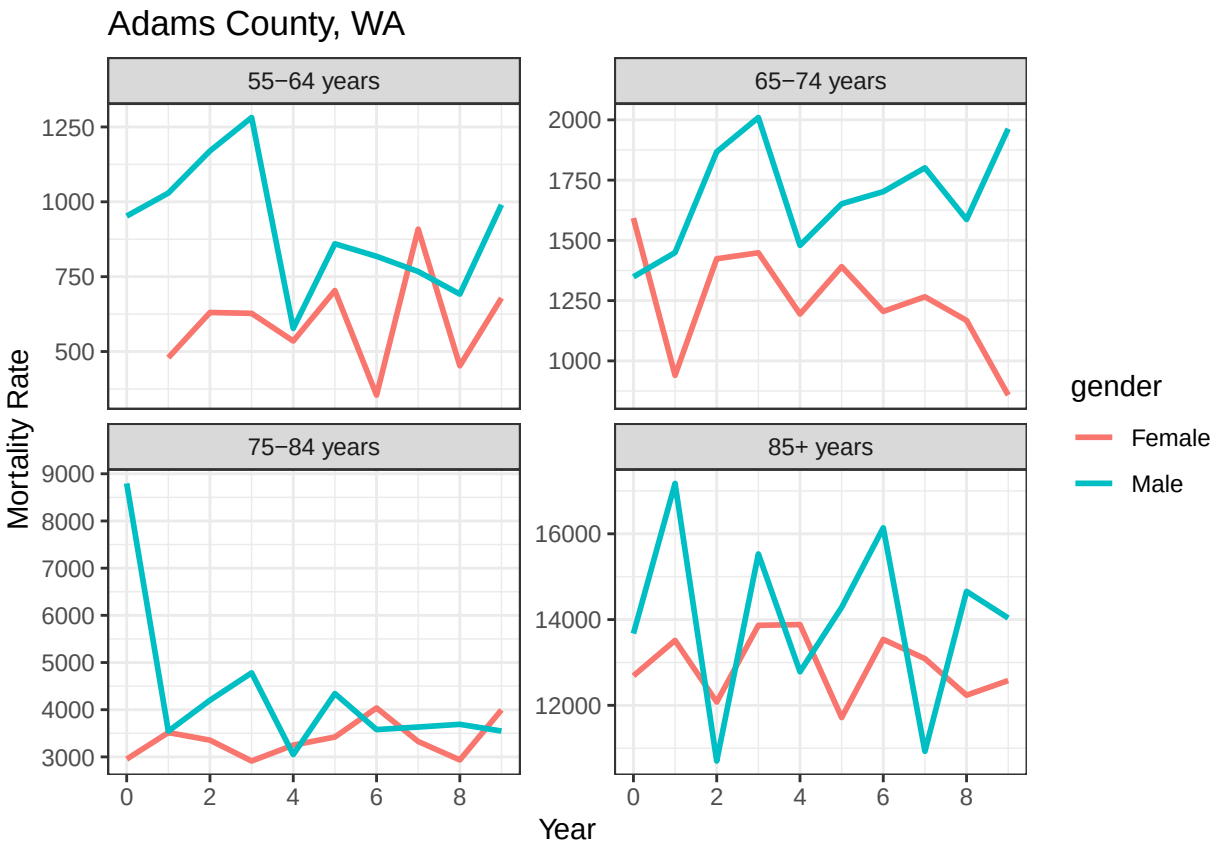

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

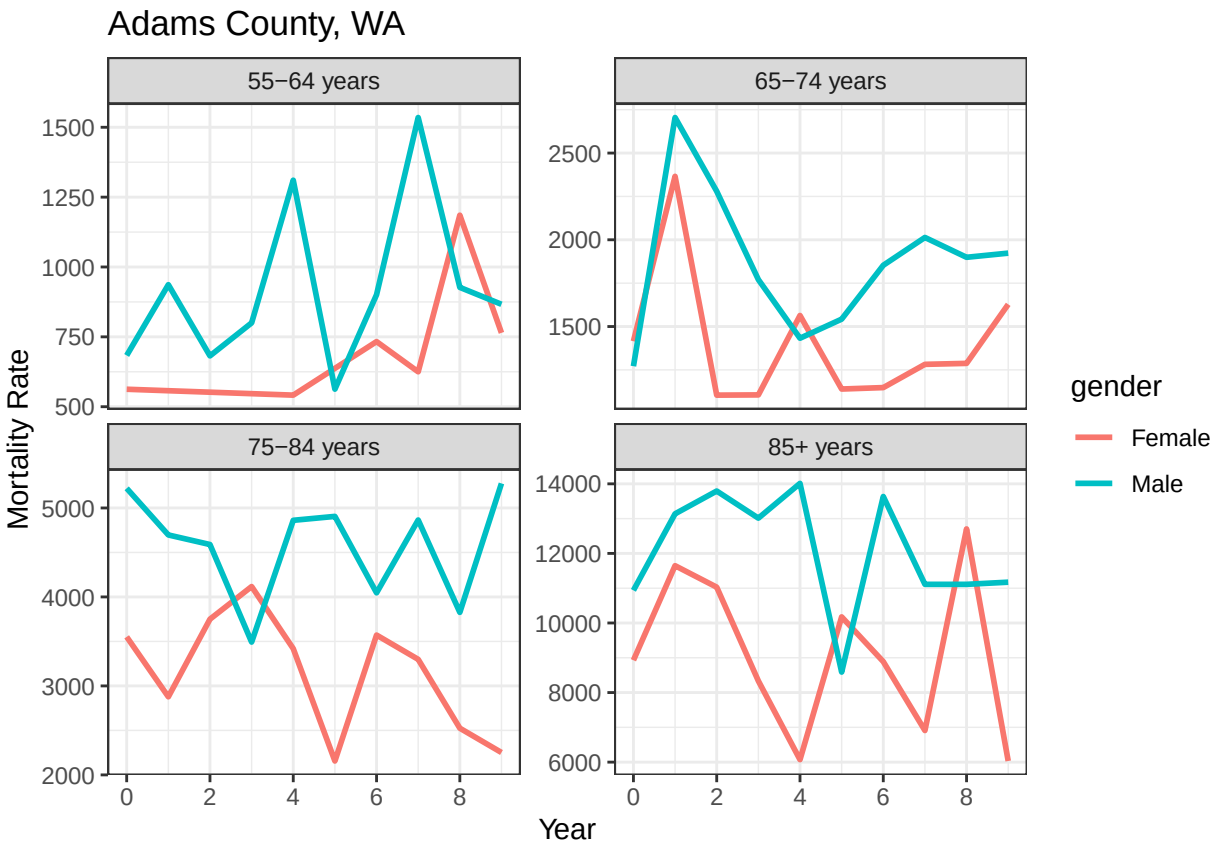

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

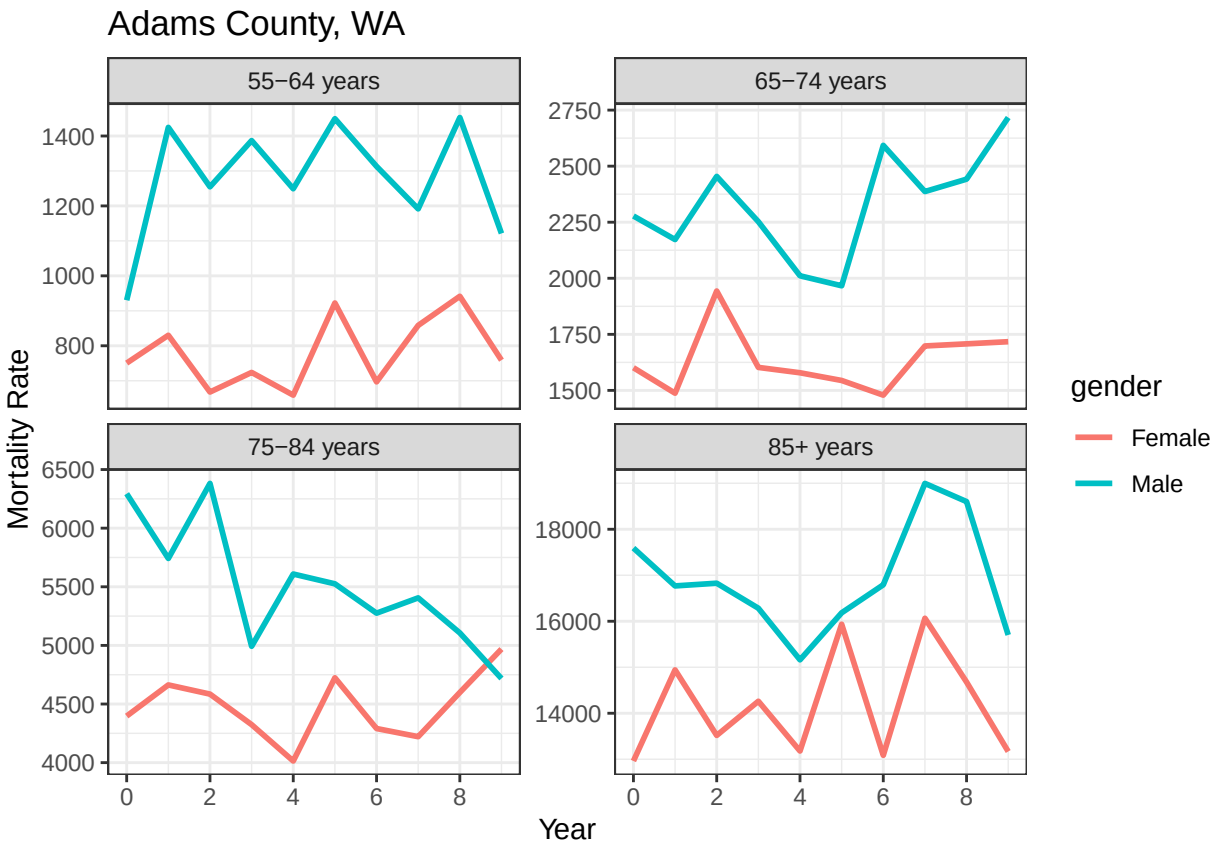

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male
900
700
500
300
1750
1500
1250
1000
750
4000
3000
16000
14000
12000
10000
0
2
4
6
8

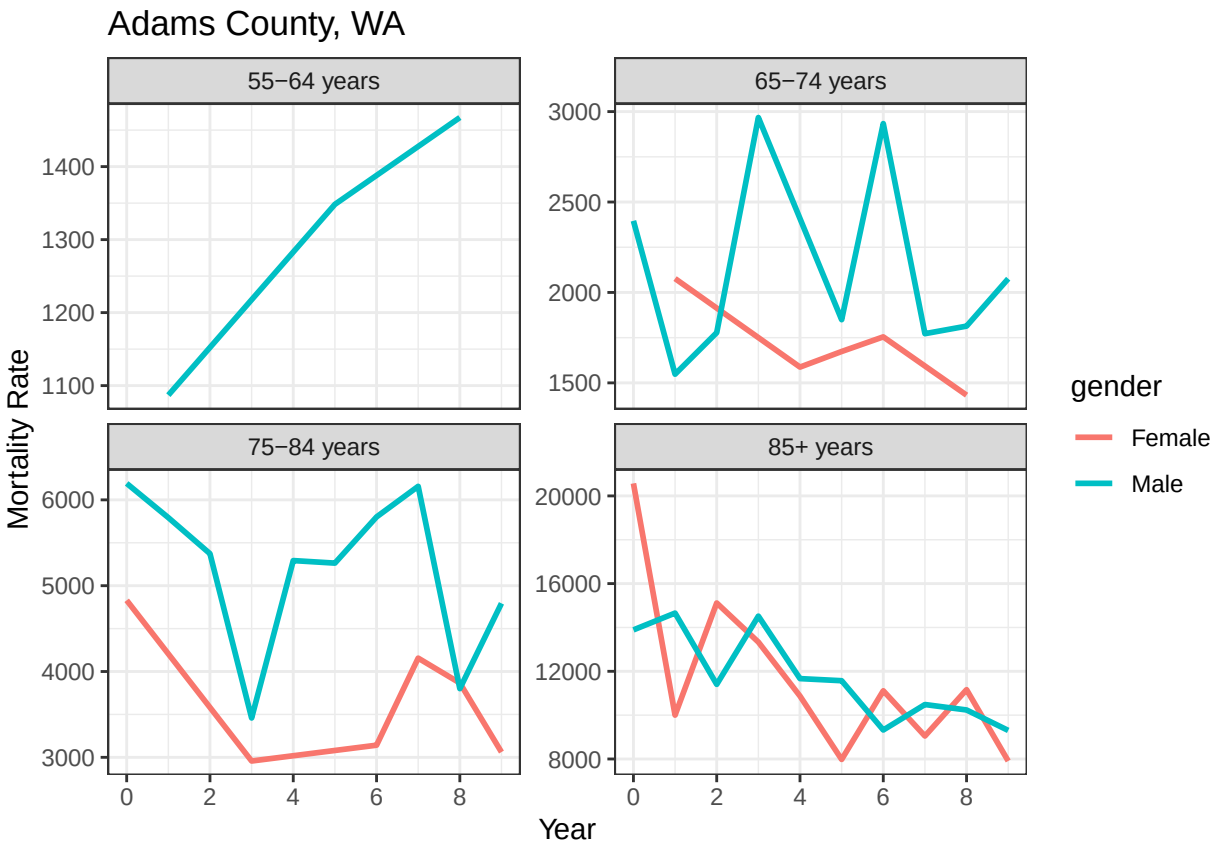

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

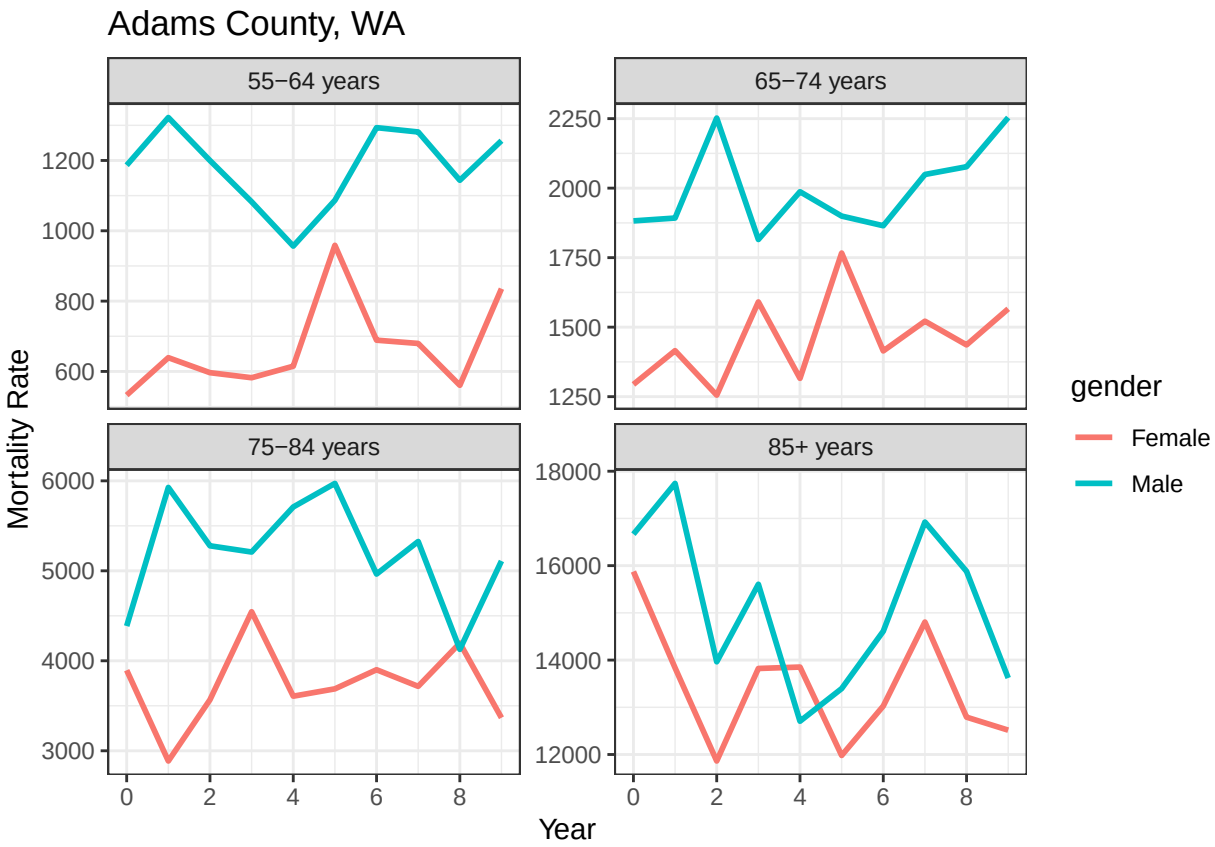

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

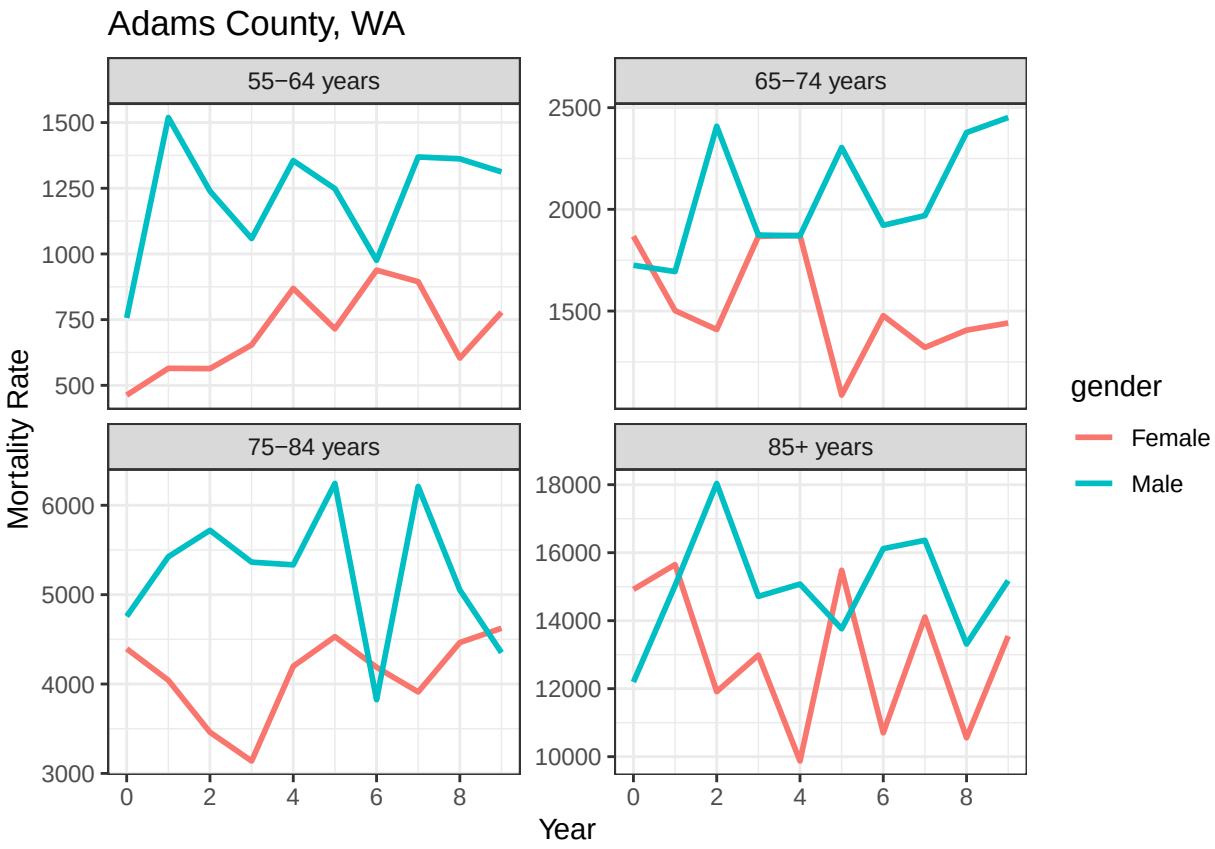

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

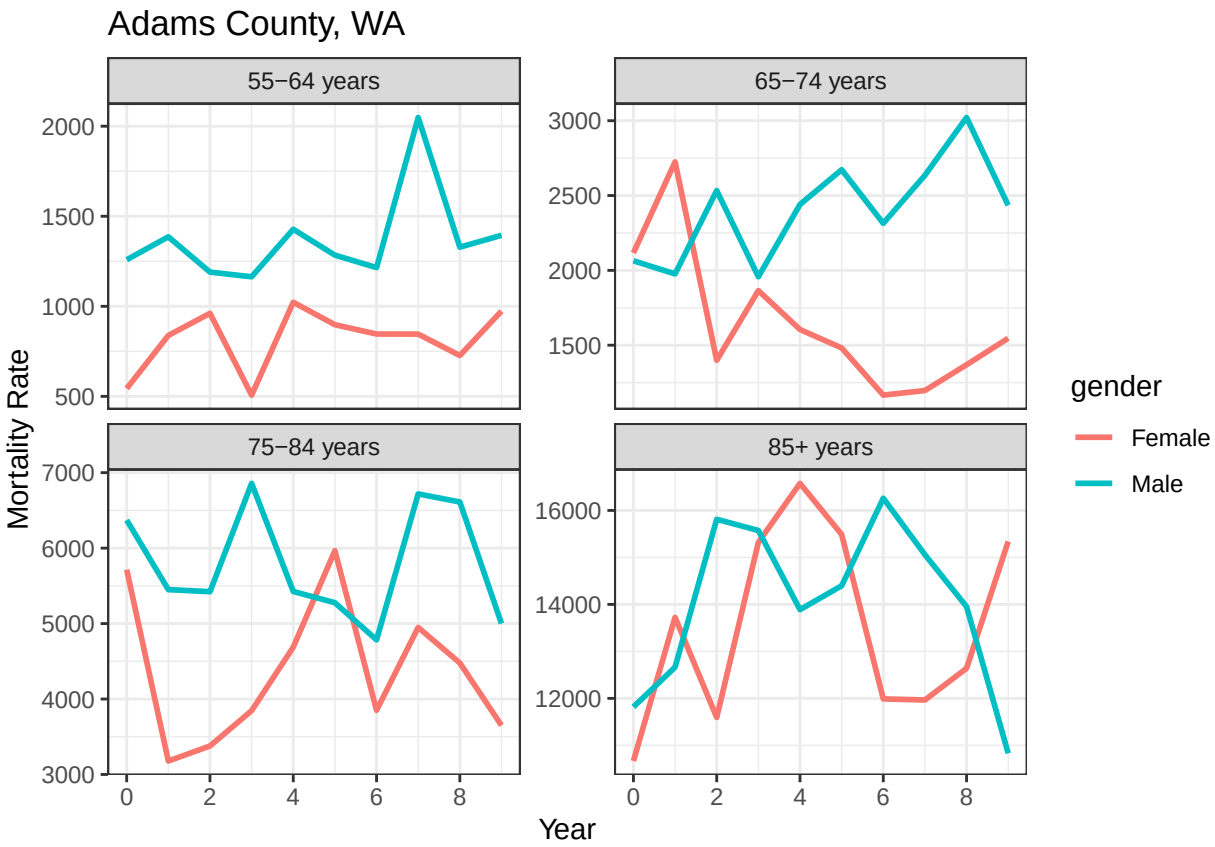

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

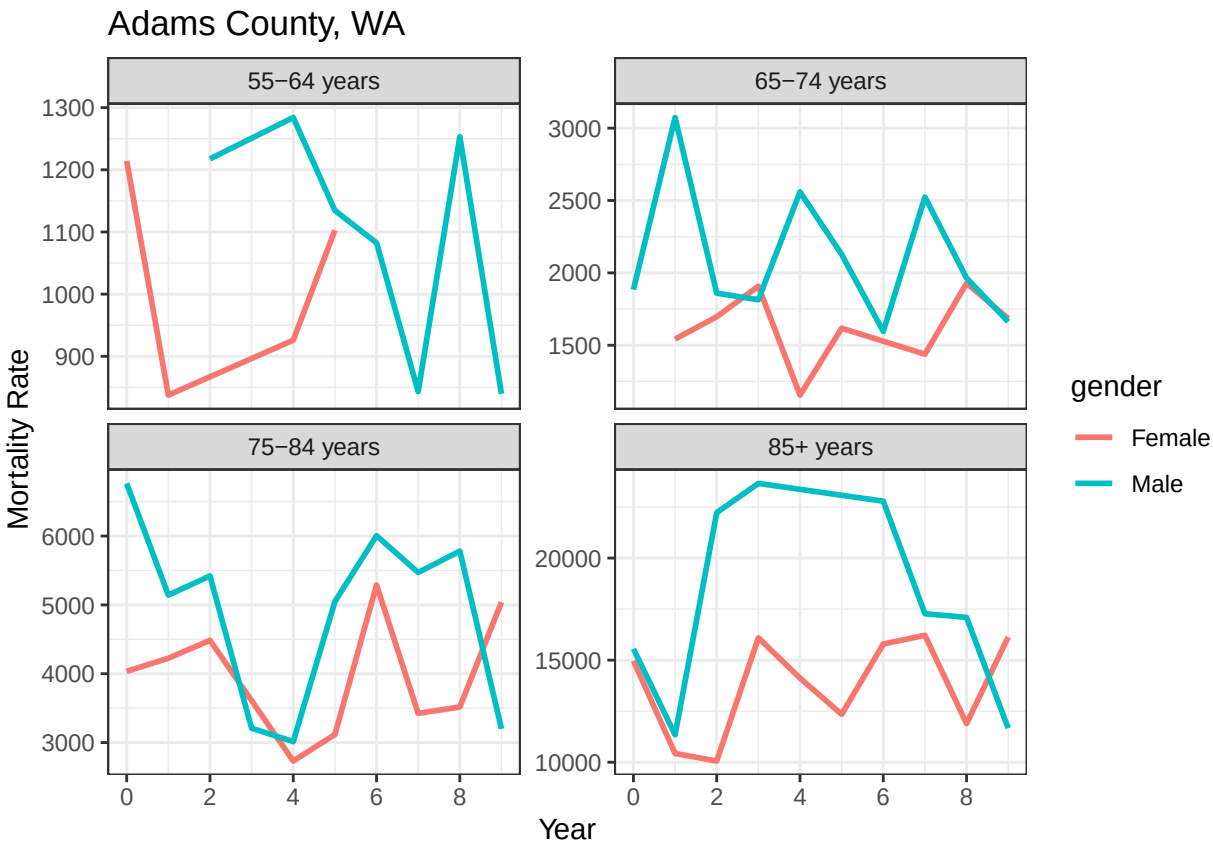

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

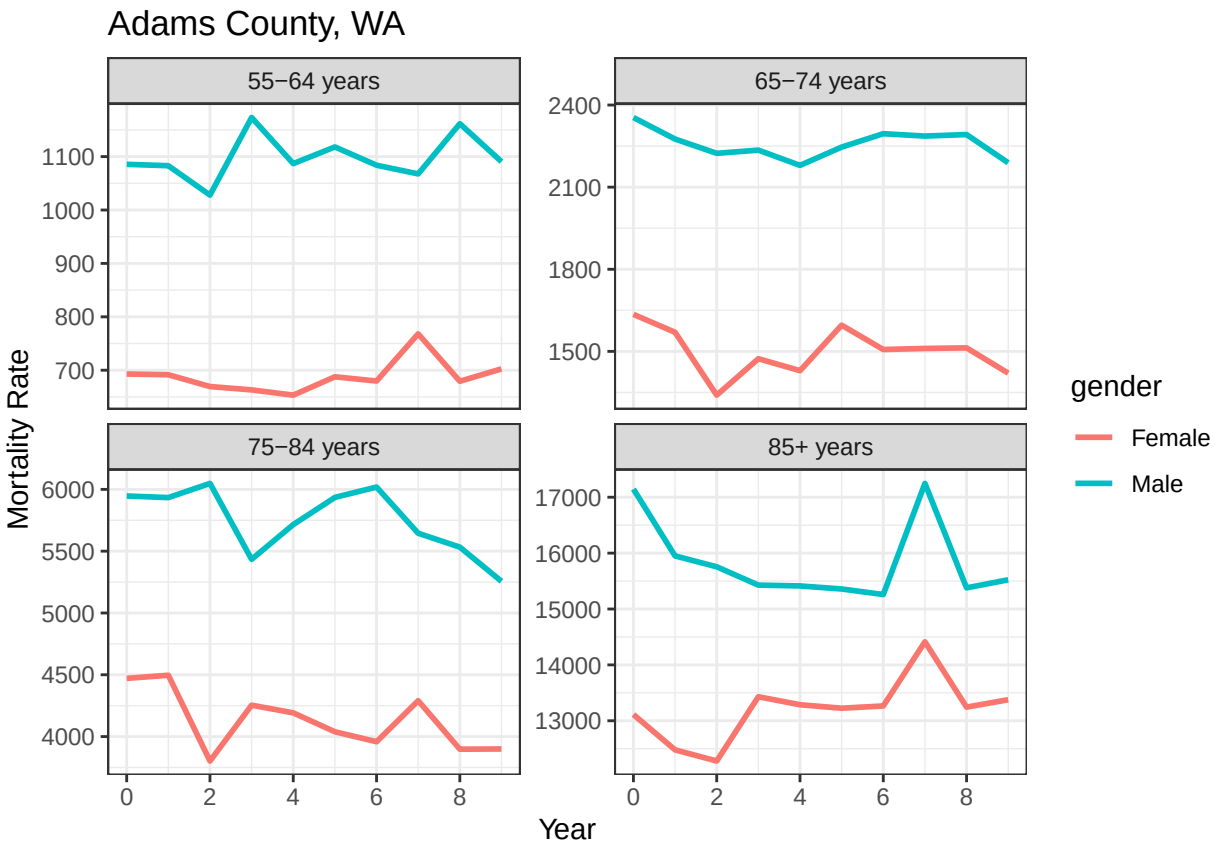

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male
1400
1300
1200
1100
3000
2500
2000
1500
6000
5000
4000
3000
20000
16000
12000
8000
0
2
4
6
8

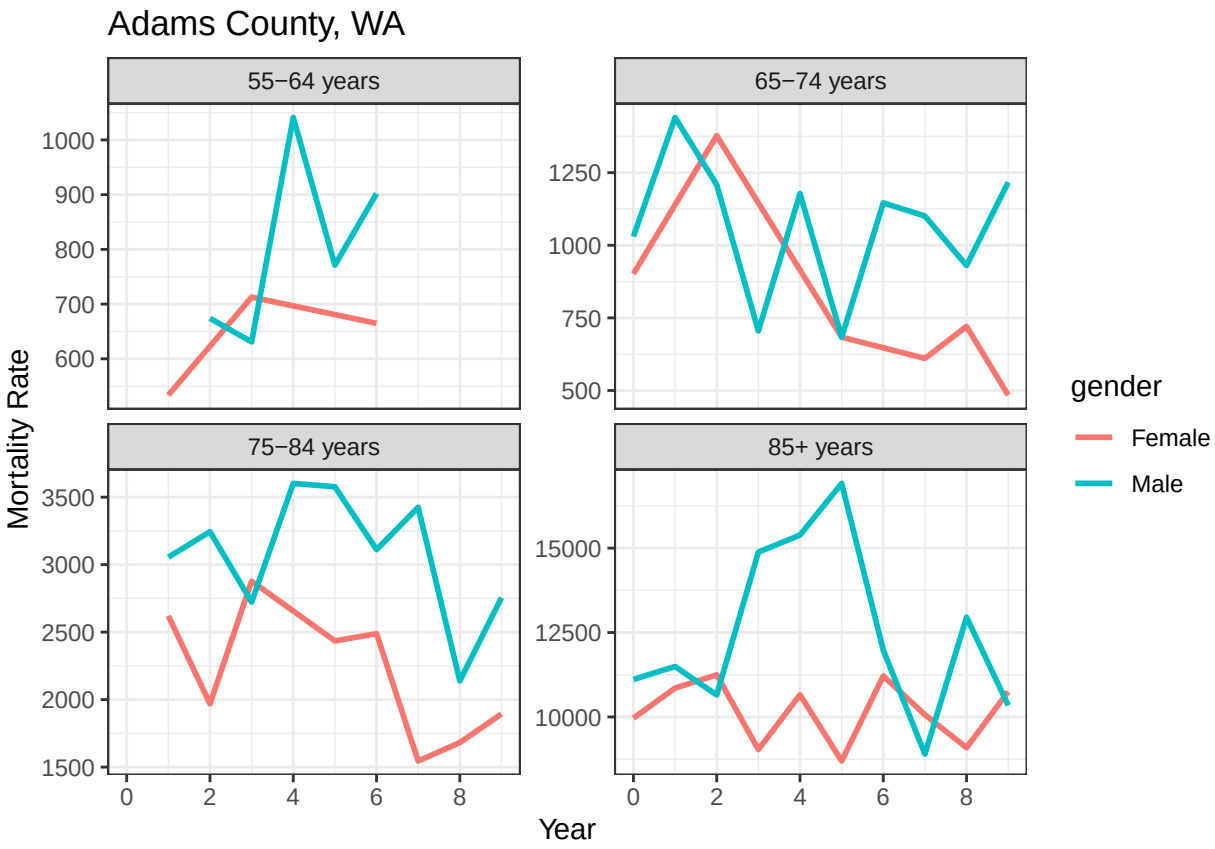

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

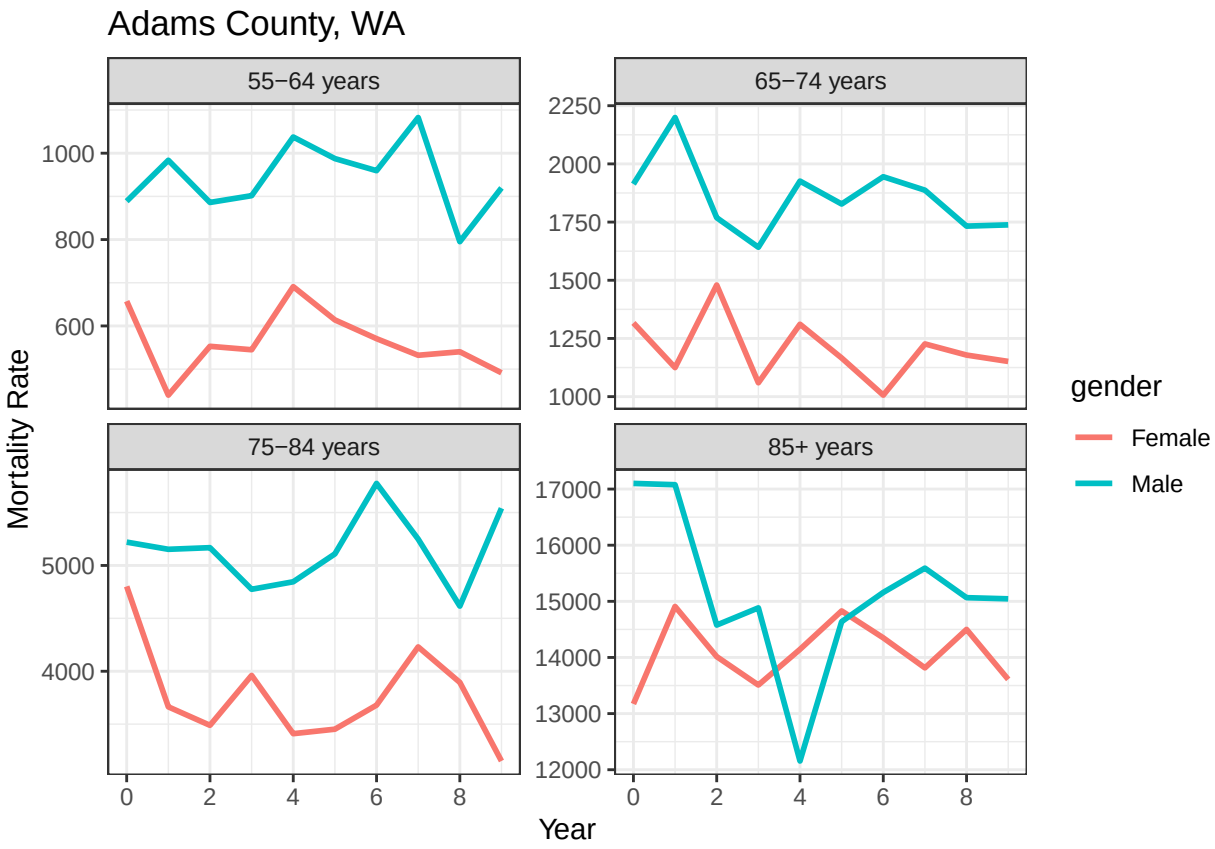

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

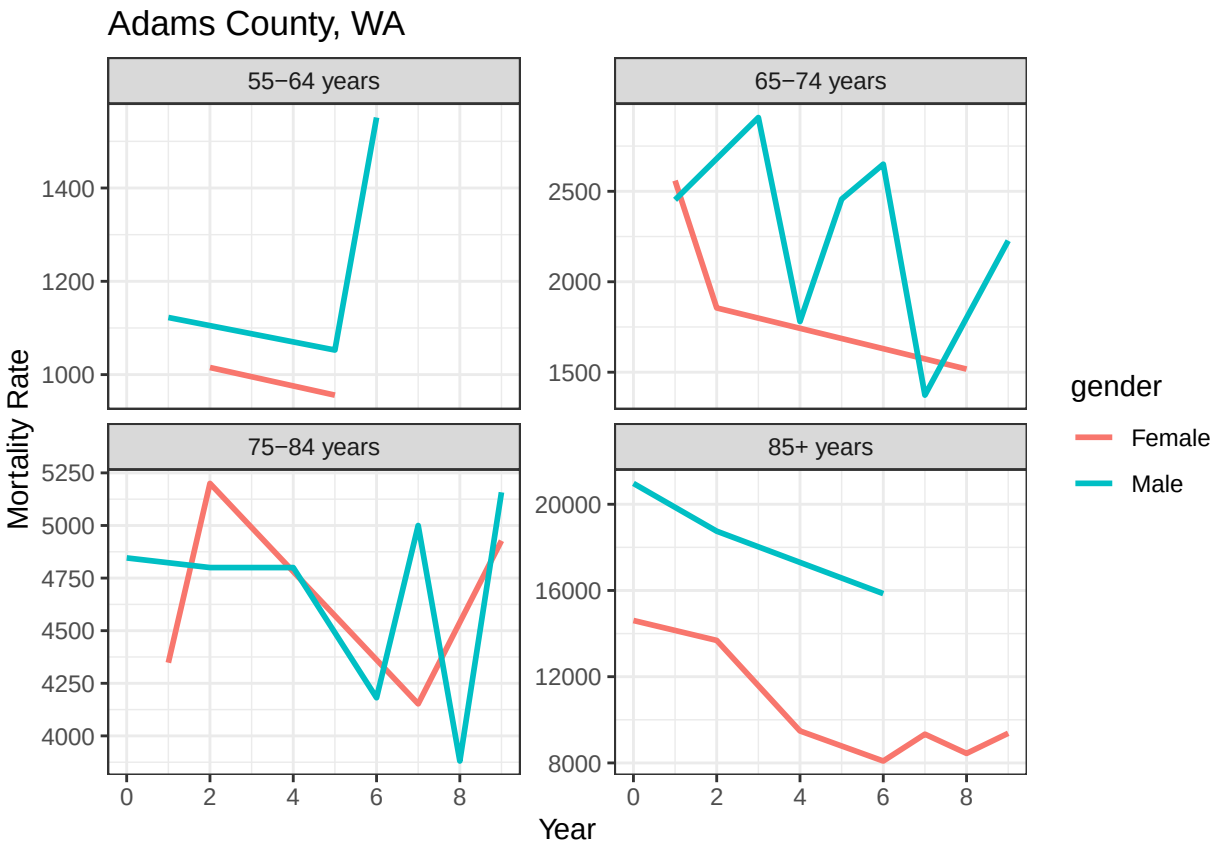

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

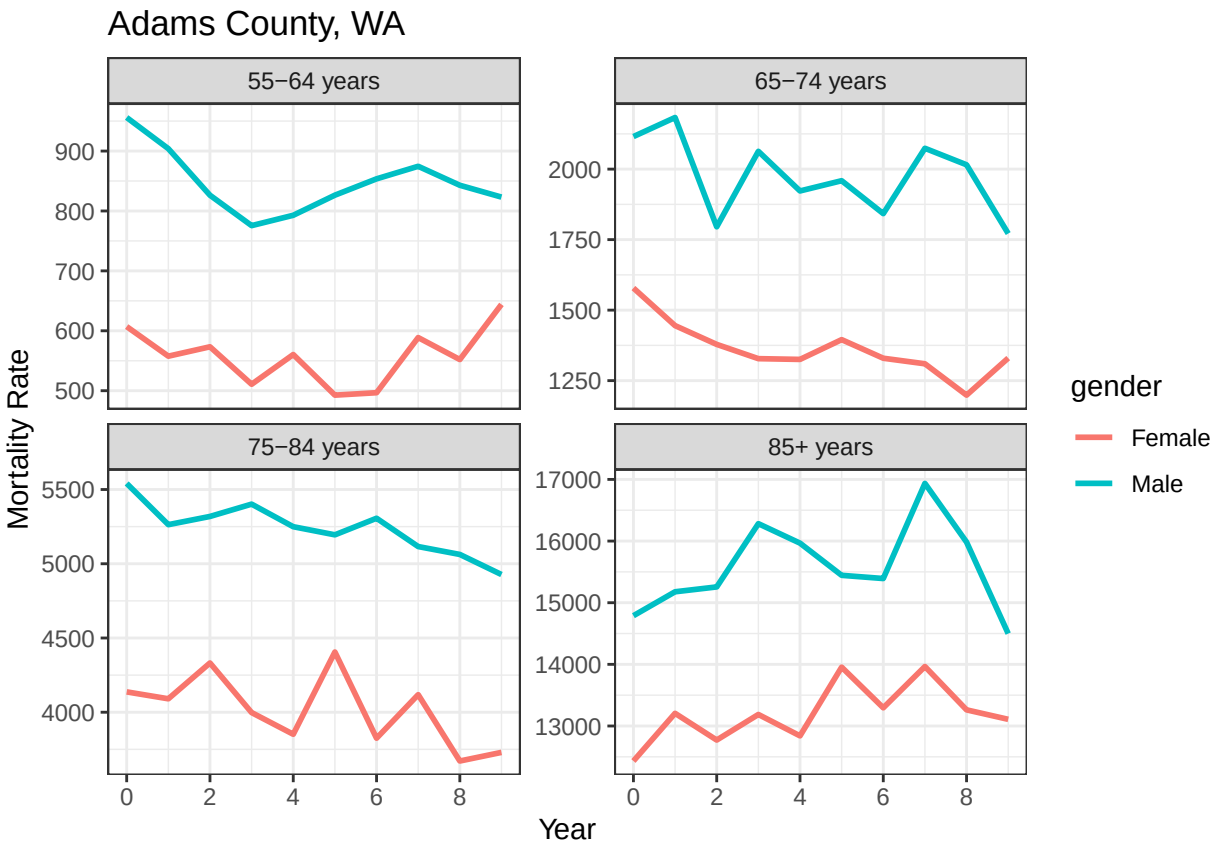

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

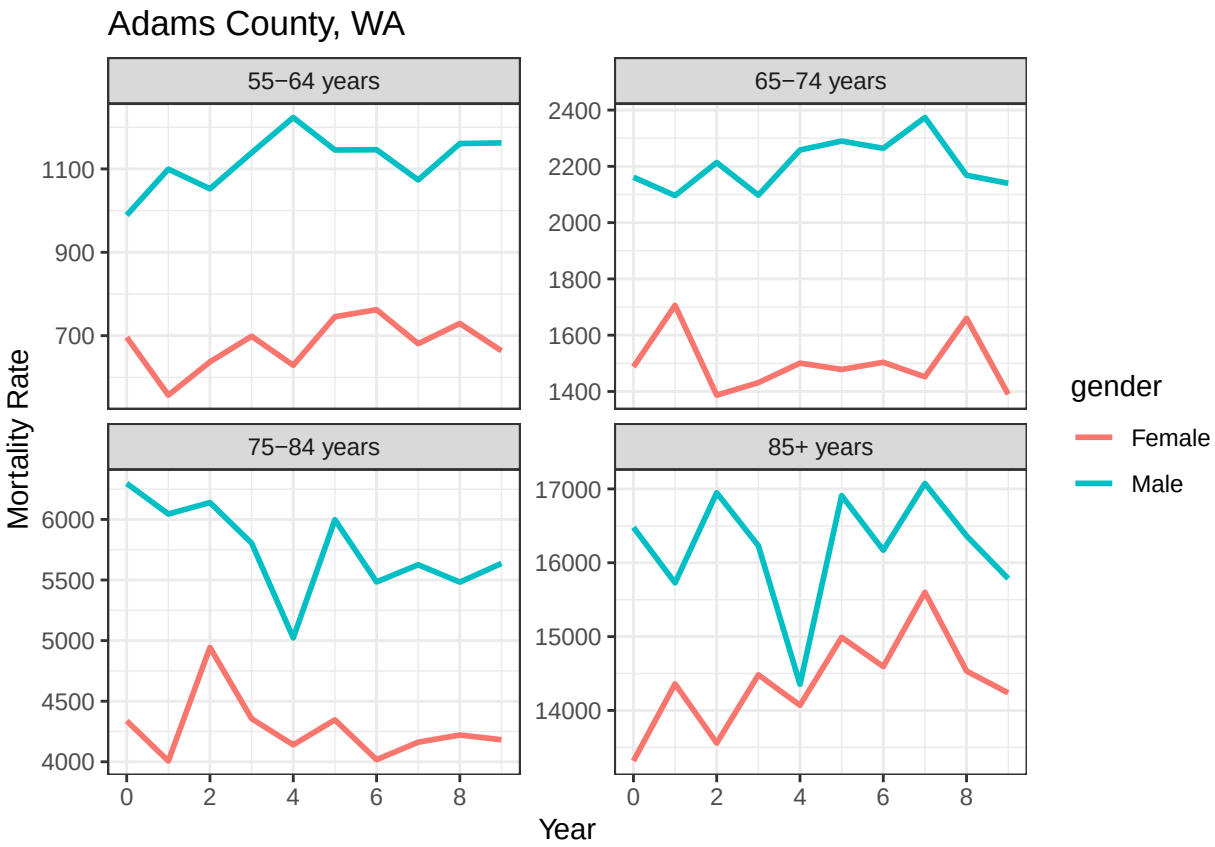

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

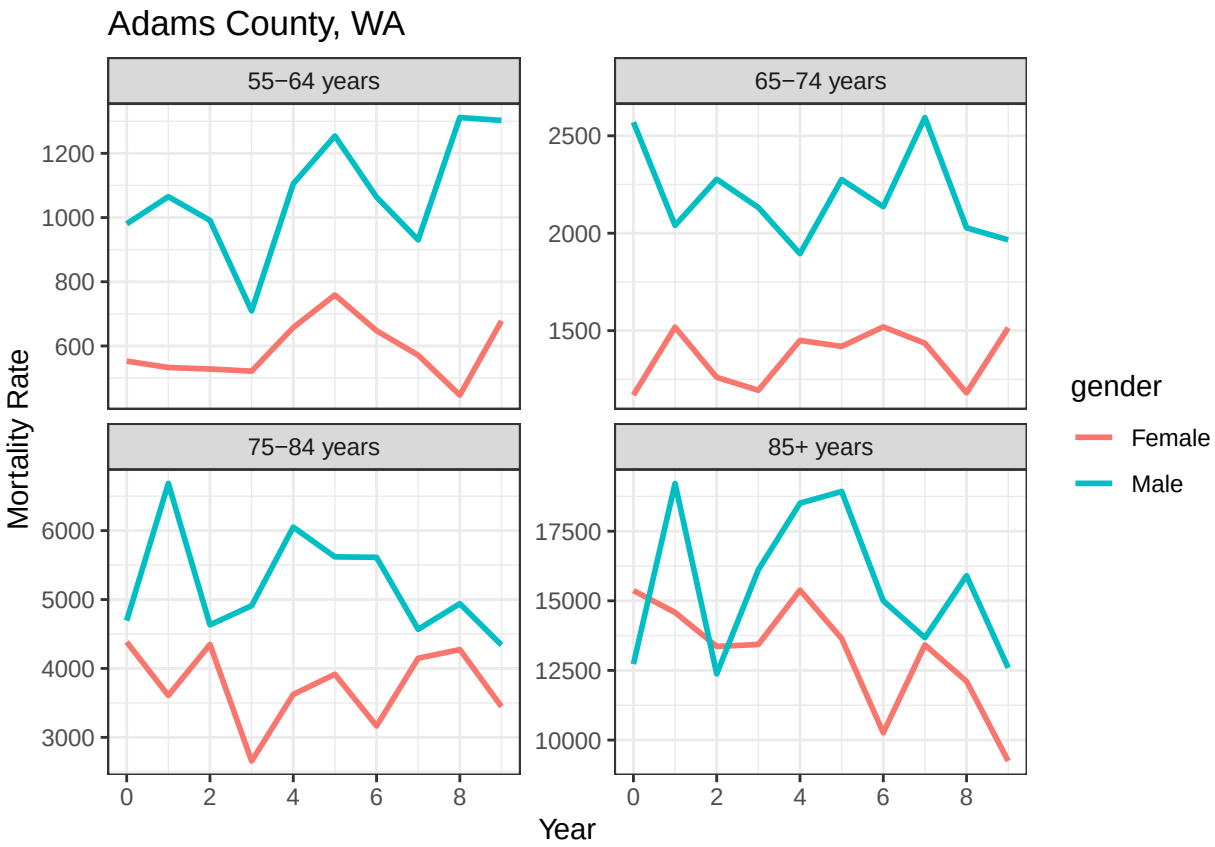

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

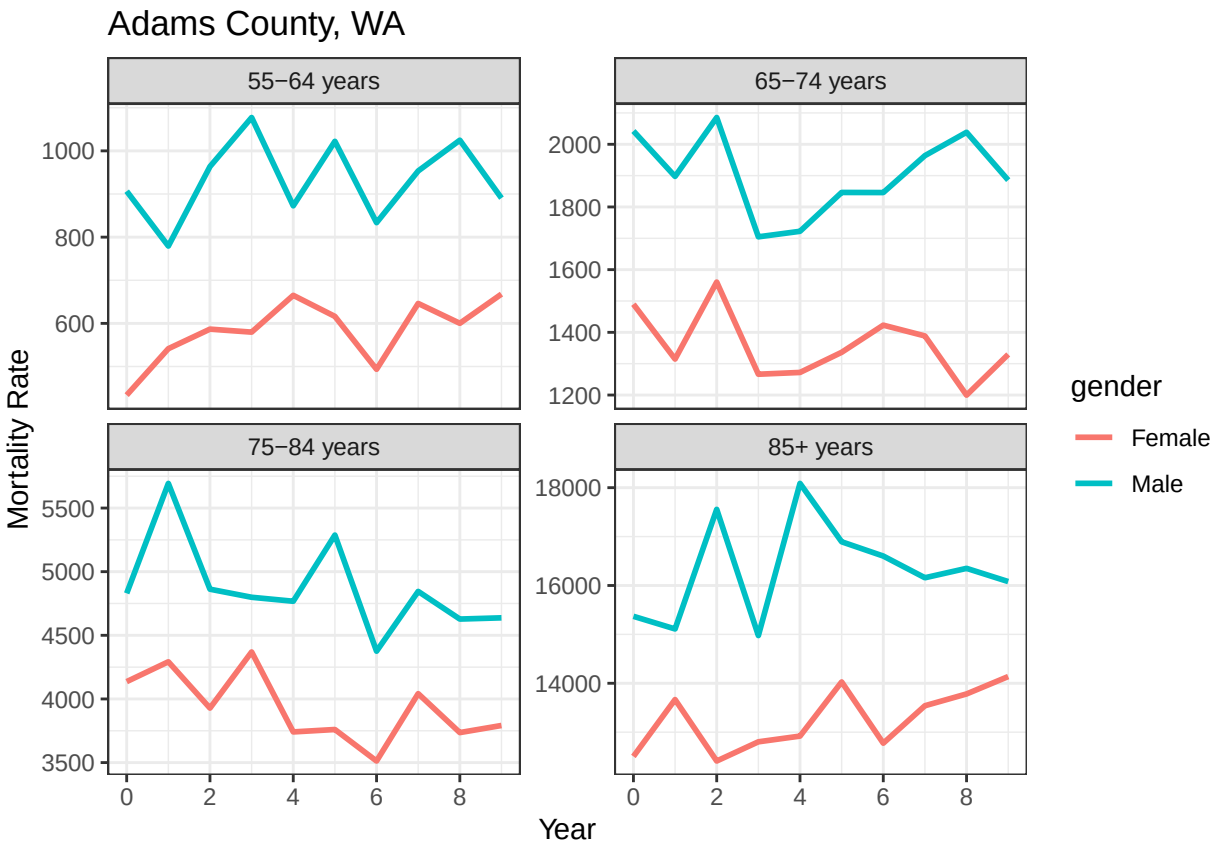

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

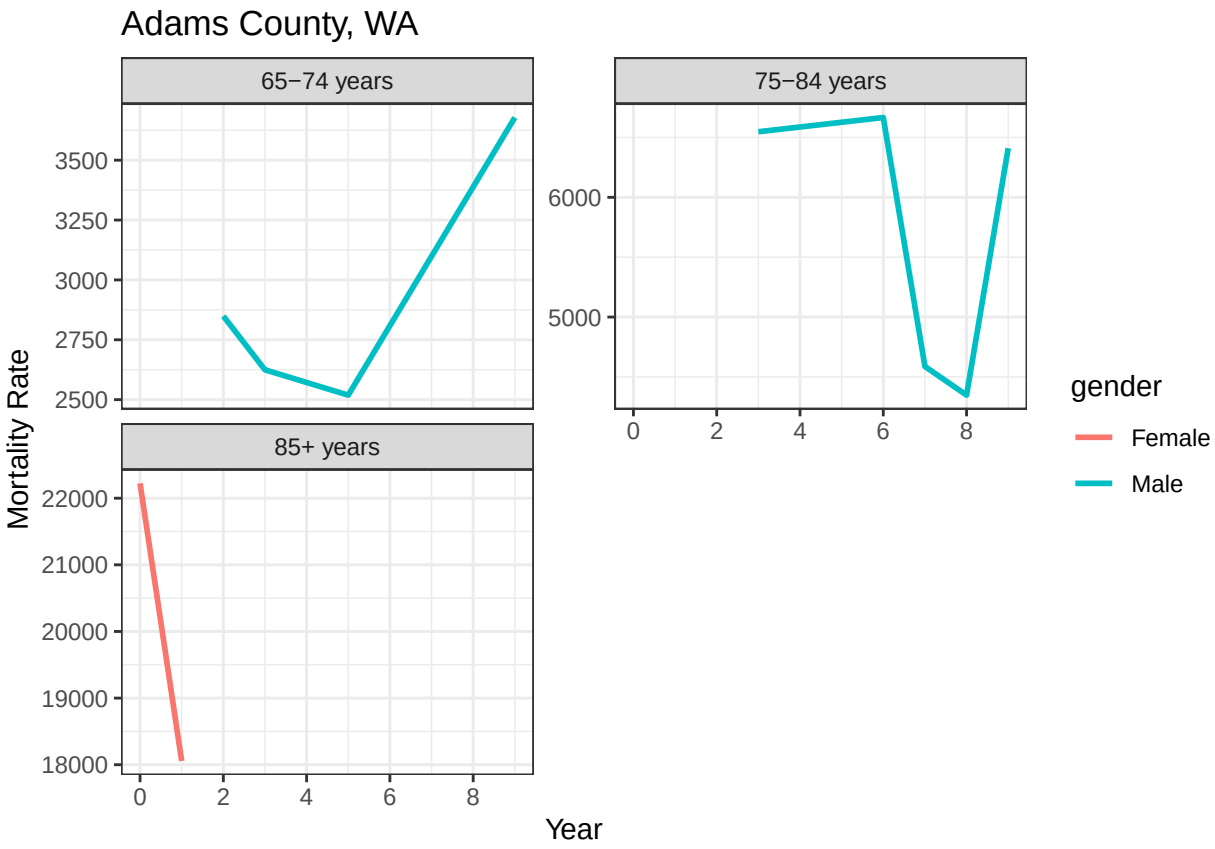

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

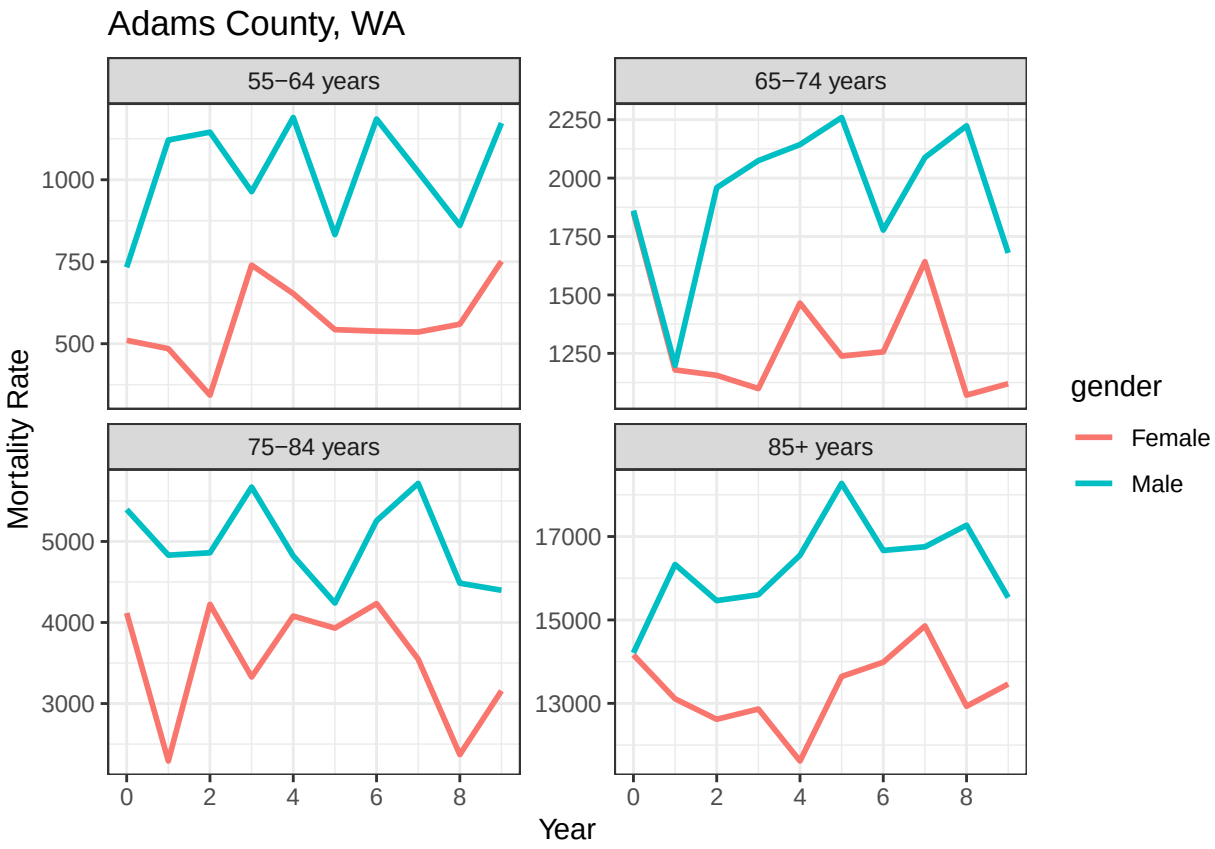

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

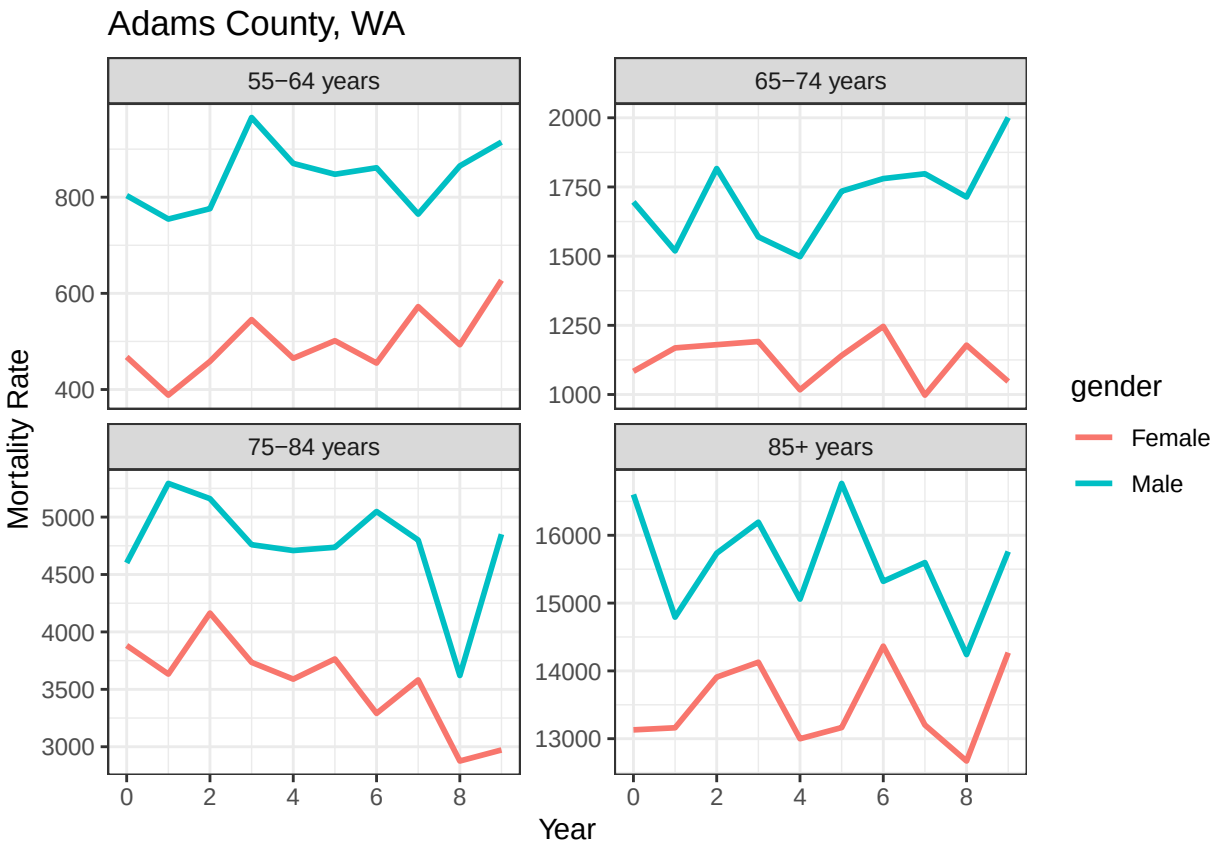

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

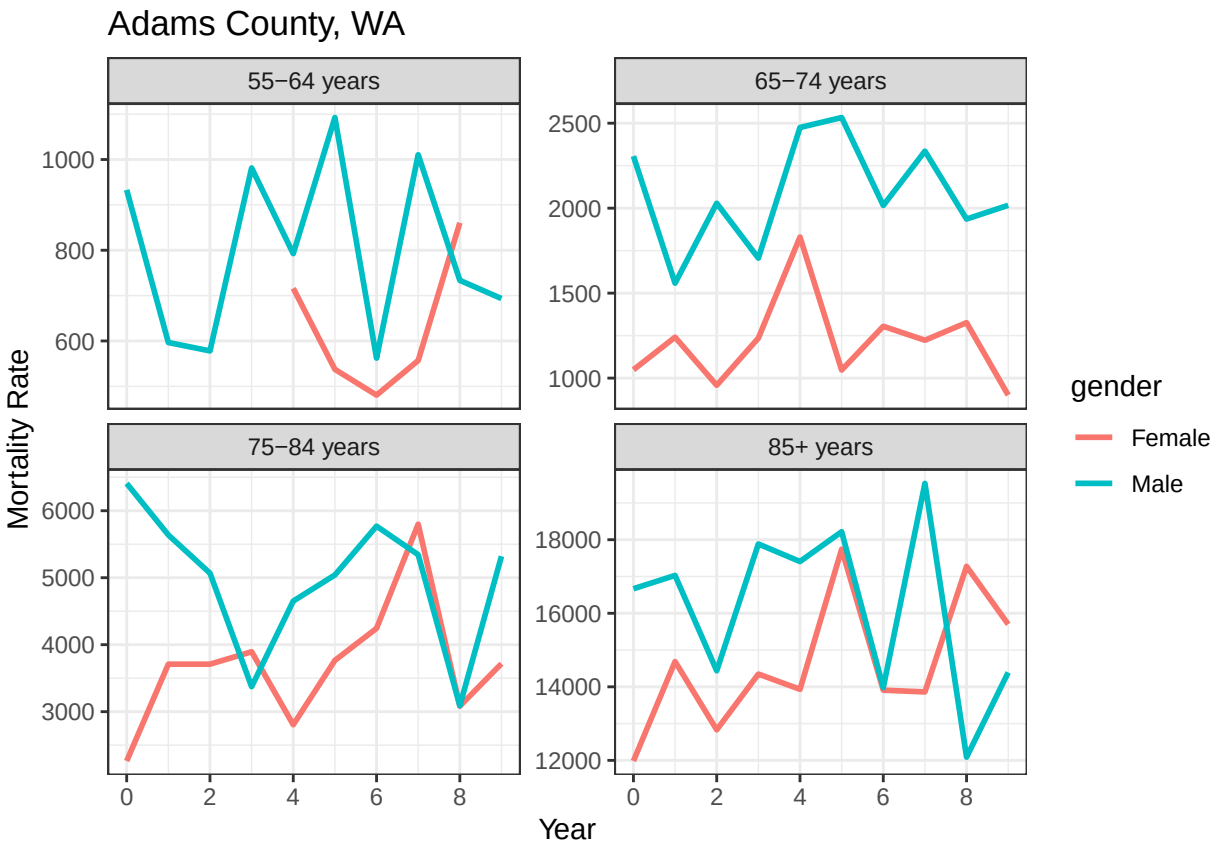

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male

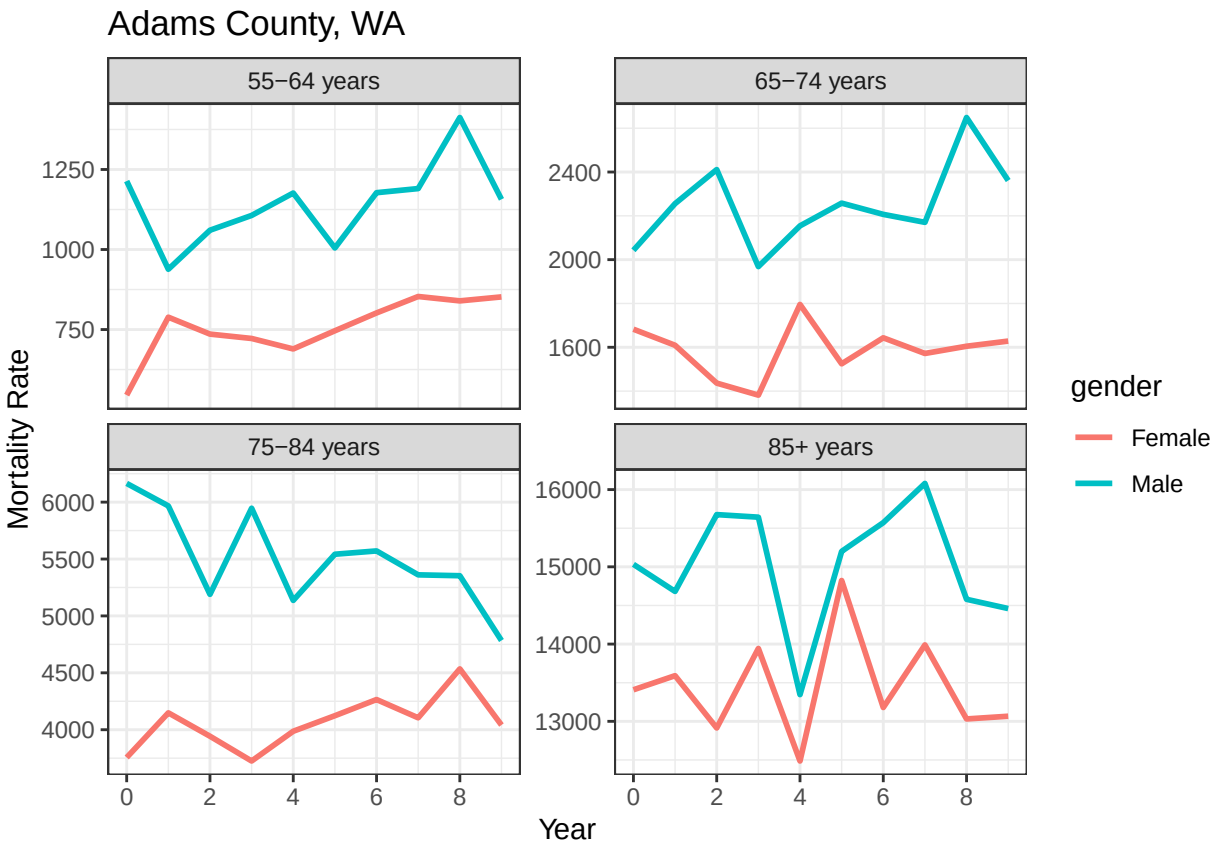

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male


Adams County, WA
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male


Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male
1000
750
500
2250
2000
1750
1500
1250
5000
4000
3000
17000
15000
13000
0
2
4
6
8

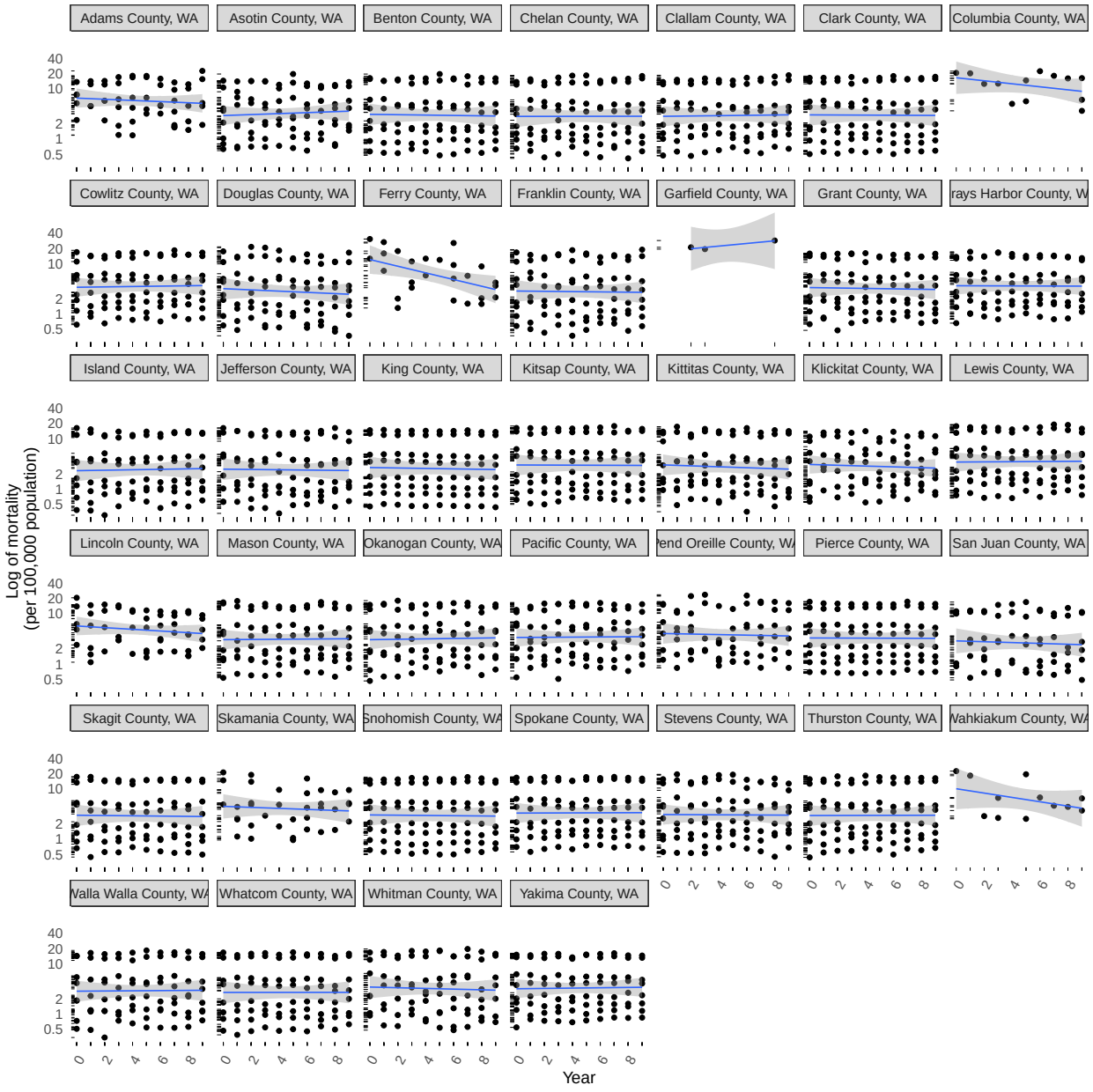

Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male


Adams County, WA
55−64 years
65−74 years
75−84 years
85+ years
Mortality Rate
Year
gender
Female
Male



Plotting mortality rate per 100,000 population across year for all the county

There seems to be a change in mortality rate across years for some counties. However, mostly this is due to data gaps. In counties where the data was more or less similar to that of state data, the previous observed trend continued.


Adams County, WA
Asotin County, WA
Benton County, WA
Chelan County, WA
Clallam County, WA
Clark County, WA
Columbia County, WA
Cowlitz County, WA
Douglas County, WA
Ferry County, WA
Franklin County, WA
Garfield County, WA
Grant County, WA
rays Harbor County, W
Island County, WA
Jefferson County, WA
King County, WA
Kitsap County, WA
Kittitas County, WA
Klickitat County, WA
Lewis County, WA
Lincoln County, WA
Mason County, WA
Okanogan County, WA
Pacific County, WA
end Oreille County, W
Pierce County, WA
San Juan County, WA
Skagit County, WA
Skamania County, WA
nohomish County, WA
Spokane County, WA
Stevens County, WA
Thurston County, WA
Vahkiakum County, WA
Walla Walla County, WA
Whatcom County, WA
Whitman County, WA
Yakima County, WA
Log of mortality
(per 100,000 population)
40
20
10
2
1
0.5
0
2
4
6
8
Year

Code Appendix

```
### Setting up the packages
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
# check if packages are installed; if not, install them
packages <- c("tidyverse", "readr", "ggExtra", "plotly",
              "ggplot2","ggstatsplot","ggside","rigr","nlme","lmtest",
              "sandwich","gtsummary")
not_installed <- setdiff(packages, rownames(installed.packages()))
if (length(not_installed)) install.packages(not_installed)

# load packages
library("MASS")
library(sandwich)
library(gtsummary)
library(readr)
library(lmtest)
library(nlme)
library(table1)
library(ggstatsplot)
library(ggside)
library(rigr)
library(ggExtra)
library(broom)
library(plotly)
library(ggplot2)
library(tidyverse) # don't load tidyverse package due to conflict with dplyr
### -------------------------------------------------------------
#Loading working directory of the raw data

#Please load your data/directory by changing it with your work directory
#Throughout this code module you will see a tone of places, where
#data is read and written, so please make sure to change them to your
#working directory folder format

working_directory_data <- setwd("C:/Users/latera/Desktop/Bio_ass")

mortality <- read_csv("Data/cdc-wonder-wa.csv")
knitr::kable(table1(~ county + gender + age + deaths + pop, data=mortality,
       overall="Total"),

      digits = 5,caption = "Summary of our data")
mortality$emprical_rate <- (mortality$deaths / mortality$pop) * 100000
n<-mortality%>%arrange(pop) %>% head(5)
knitr::kable(n,
             digits = 5,caption = "The first five observations ordered by population")

xbreaks <- c(30, 100, 500, 1000, 2000, 5000,
10000, 20000, 50000,100000, 200000)
p <- ggplot(mortality, aes(x=pop, y=deaths, col=county, shape=age)) +
      geom_point(size=1.2)+
      xlab("Body Mass Index (BMI)") + ylab("Charges \n (dollars)")+
```

```
  scale_x_log10(breaks=xbreaks,
labels=xbreaks/1000)+
  geom_rug(col="black",linewidth=0.20)+
  theme_bw() +
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())+
  guides(color = "none")+
  theme(legend.background = element_rect
        (fill = "transparent"))

p
#Plotting scatter plot between weight and height (inches)
library(scales)
xbreaks <- c(10, 100, 500, 1000, 2000, 10000, 20000, 40000)
p <- ggplot(mortality, aes(x= year, y=emprical_rate, col=age)) +
      geom_point(size=1.2)+
      xlab("Year") + ylab("Log of mortality \n(per 100,000 population)")+
  geom_smooth(method = "lm", size=0.445)+
  scale_y_log10(breaks=xbreaks,
labels=xbreaks/1000)+
  scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))+
  facet_wrap(~gender)+
  geom_rug(col="black",linewidth=0.20)+
  theme_bw() +
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())+
  theme(legend.background = element_rect
        (fill = "transparent"))+
  theme(axis.text.x=element_text(angle=60, hjust=1))

p
#Plotting scatter plot between weight and height (inches)
xbreaks <- c(10, 100, 500, 1000, 2000, 4000)
p <- ggplot(mortality, aes(x=age, y=emprical_rate, col=age)) +
      geom_boxplot()+
      xlab("Age group") + ylab("Log of mortality rate \n per 100,000 thousand population")+
  geom_jitter(alpha=0.1)+
  scale_y_log10(breaks=xbreaks,
labels=xbreaks/1000)+
  geom_rug(col="black",linewidth=0.20)+
  theme_bw() +
  facet_wrap(~gender)+
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
    panel.grid.major = element_blank(),
```

```
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())+
  theme(legend.background = element_rect
        (fill = "transparent"))

p+ theme(legend.position = "none")

#Plotting scatter plot between weight and height (inches)
library(scales)
xbreaks <- c(10, 100, 500, 1000, 2000, 10000, 20000, 40000)
p <- ggplot(mortality, aes(x= year, y=emprical_rate)) +
      geom_point(size=1.2)+
      xlab("Year") + ylab("Log of mortality \n(per 100,000 population)")+
  geom_smooth(method = "lm", size=0.445)+
  scale_y_log10(breaks=xbreaks,
labels=xbreaks/1000)+
scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))+
  geom_rug(col="black",linewidth=0.20)+
  theme_bw() +
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())+
  theme(legend.background = element_rect
        (fill = "transparent"))+
  theme(axis.text.x=element_text(angle=60, hjust=1))

p + theme(legend.position = "none")

mortality$year <- mortality$year - 2010

# Fit a Poisson GLM with log link and offset for population size
model <- glm(deaths ~  gender + age + year + offset(log(pop)),
             data = mortality, family = poisson(link = "log"))

model %>%
  tbl_regression(
    label = list(

      gender ~ "Gender",
      age ~ "Age",
      year ~ "Year"
    ),
    exponentiate = FALSE,
    intercept = TRUE,
    estimate_fun = purrr::partial(style_number, digits = 5)
    )


model <- glm(deaths ~  gender + age + year + year:age + offset(log(pop)),
```

```
            data = mortality, family = poisson(link = "log"))

model %>%
  tbl_regression(
    label = list(

      gender ~ "Gender",
      age ~ "Age",
      year ~ "Year"
    ),
    exponentiate = FALSE,
    intercept = TRUE,
    estimate_fun = purrr::partial(style_number, digits = 5)
    )


# list of counties
counties <- unique(mortality$county)

# loop through each county and create a plot
for (c in counties) {
  # filter data for the county
  df_county <- mortality[mortality$county == c,]
  # Create the plot
p<-ggplot(df_county, aes(x = year, y = emprical_rate, group = interaction(gender, age), color = gender))
  geom_line(size = 1) +
  facet_wrap(~ age, ncol = 2, scales = "free_y") +
  scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))+
  xlab("Year") +
  ylab("Mortality Rate") +
  ggtitle(mortality$county)+

  theme_bw()
print(p)



}

library(scales)
xbreaks <- c(10, 100, 500, 1000, 2000, 10000, 20000, 40000)
p <- ggplot(mortality, aes(x= year, y=emprical_rate)) +
      geom_point(size=1.2)+
      xlab("Year") + ylab("Log of mortality \n(per 100,000 population)")+
  geom_smooth(method = "lm", size=0.445)+
  scale_y_log10(breaks=xbreaks,
labels=xbreaks/1000)+
facet_wrap(~county)+
  geom_rug(col="black",linewidth=0.20)+
  theme_bw() +
  scale_x_continuous(breaks = ~ axisTicks(., log = FALSE))+
  theme(axis.line = element_line(colour = "white"),
        axis.ticks = element_blank(),
```

```
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank())+
  theme(legend.background = element_rect
        (fill = "transparent"))+
  theme(axis.text.x=element_text(angle=60, hjust=1))

p + theme(legend.position = "none")
```