Homework Three [76 pts]

PA 541 - Spring 2021

3/22/2021

PART ONE

Load the data file called 'cavax.csv'. This data contains a school-level table of rates of Personal Belief Exemptions (PBEs) in California kindergartens for the 2014-15 school year. At that time, a PBE allowed a child to enter kindergarten without having received the usual complement of vaccinations. The data contains the following variables:

Variable name (Description)

- *code* (School)
- *county* (County name)
- *name* (School name)
- *type* (School type (public or private))
- *district* (School district name)
- *city* (City location)
- *enrollment* (Total kindergarten enrollment)
- *pbe_pct* (Rounded percent of personal belief exemptions)
- *exempt* (Percent of students exempt from providing vaccination records)
- *med_exempt* (Percent exempt due to medical reasons)
- *rel_exempt* (Percent exempt due to religious reasons)

QUESTION 1 [8 pts]

a. In how many schools is the percentage of students exempt for medical reasons (med_exempt) greater than the percentage exempt for religious reasons (rel_exempt)? [2 pts] Of this set of schools, what percent are public schools? [2 pts]

```
cavax$question1 <- ifelse(cavax$med_exempt > cavax$rel_exempt, 1, 0)
table(cavax$question1)
```

```
##
##    0    1
## 6514  518
```

```
cavax1 = cavax %>%
  filter(question1 == 1)

round(prop.table(table(cavax1$type)),2)
```

```
##
## PRIVATE  PUBLIC
##    0.14    0.86
```

b. Which county, when averaging across all the schools in that county, has the highest average percentage of exempt students (exempt)? Note, we are using the variable exempt here. [2 pts]

```
cavax2 = cavax %>%
  group_by(county) %>%
  summarize(avg_exempt = mean(exempt)) %>%
  arrange(desc(avg_exempt))
```

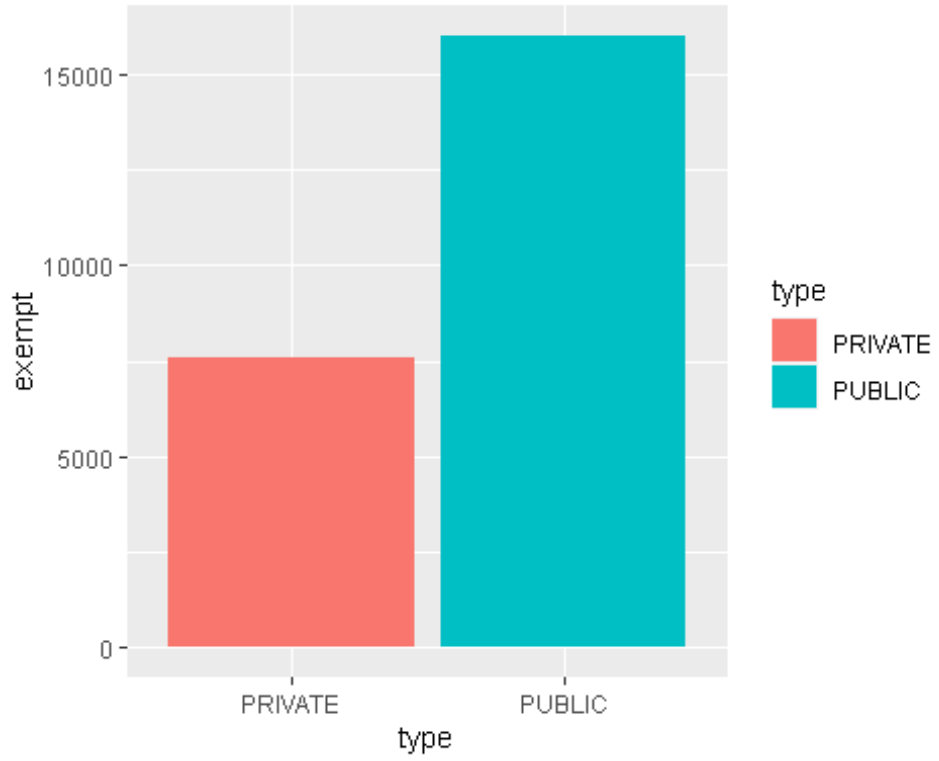
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
cavax2
```

```
## # A tibble: 57 x 2
##    county     avg_exempt
##    <chr>           <dbl>
##  1 NEVADA          24.0
##  2 MARIPOSA        16.5
##  3 HUMBOLDT        14.2
##  4 DEL NORTE       13.2
##  5 LASSEN          12.5
##  6 SANTA CRUZ      11.8
##  7 EL DORADO       10.6
##  8 CALAVERAS        9.80
##  9 SHASTA           9.78
## 10 MENDOCINO        9.37
## # ... with 47 more rows
```

c.Create a bar chart that shows for private and public schools (type) the percent of students exempt from providing vaccination records (exempt). [2 pts]

```
ggplot(cavax, aes(x=type, y=exempt, fill = type)) + geom_col()
```



QUESTION 2 [10 pts]

Estimate a model predicting exempt (exempt) by district type (type) and enrollment (enrollment). [2 pts] Interpret the intercept and the coefficients. [4 pts] What is the predicted exempt percentage for a public school with 100 students in kindergarten? [2 pts] What is the predicted exempt percentage for a private school with 80 students in kindergarten? [2pts]

```
mod1 <- lm(exempt ~ type + enrollment, data = cavax)
summary(mod1)
```

```
##
## Call:
## lm(formula = exempt ~ type + enrollment, data = cavax)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.993 -3.296 -1.727  0.588 86.986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.288602   0.209140  30.069  < 2e-16 ***
## typePUBLIC  -0.860822   0.263483  -3.267  0.00109 **
## enrollment  -0.029606   0.002383 -12.425  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.396 on 7029 degrees of freedom
```

```
## Multiple R-squared:  0.04034,    Adjusted R-squared:  0.04007
## F-statistic: 147.7 on 2 and 7029 DF,  p-value: < 2.2e-16

##Intercept:a private school with 0 enrollment is predicted to have 6.3% exem
pt
##type: the reference group is private schools. The coefficient for public sc
hools is -0.86 meaning there are 0.86 percent points fewer exempt students at
public schools compared to private schools. The relationship is significant (
p<.01).
##enrollment: the coefficient for enrollment is -0.03 meaning that for each a
dditional student enrolled in kindergarten, the percentage of students exempt
from providing vaccination records decreases by 0.03 percentage points. The r
elationship is significant (p<.001).

##public school with 100 students in kindergarten
y1 = 6.288602 - 0.860822 - (0.029606 * 100)
y1

## [1] 2.46718

##a different way
p1 = c(1,1,100)
crossprod(p1,coef(mod1))

##          [,1]
## [1,] 2.467164

##a private school with 80 students
y2 = 6.288602 - (0.029606 * 80)
y2

## [1] 3.920122

##a different way
p2 = c(1,0,80)
crossprod(p2,coef(mod1))

##         [,1]
## [1,] 3.92011
```

QUESTION 3 [8 pts]

Test whether the assumption of homoskedasticy has been met. [2 pts] Discuss results. [2 pts] Calculate the VIF for each variable. [2 pts] Should we be concerned with multicollinearity? [2 pts] (Note, the vif() command is in the 'car' package. You can also calculate the VIF yourself as we did in class)

```
## homeskedasticity
bptest(mod1)

##
##  studentized Breusch-Pagan test
```

```
##
## data:  mod1
## BP = 106.31, df = 2, p-value < 2.2e-16
```

```
## A different way to test ("by hand method")
cavax$resids = resid(mod1)
cavax$residssq = cavax$resids^2

mod2 <- lm(residssq ~ type + enrollment, data = cavax)
summary(mod2)
```

```
##
## Call:
## lm(formula = residssq ~ type + enrollment, data = cavax)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -125.9  -57.9  -37.1  -17.8 7439.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 136.5528     9.1287  14.959  < 2e-16 ***
## typePUBLIC  -39.2252    11.5007  -3.411 0.000652 ***
## enrollment   -0.6662     0.1040  -6.406 1.59e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 322.8 on 7029 degrees of freedom
## Multiple R-squared:  0.01512,    Adjusted R-squared:  0.01484
## F-statistic: 53.95 on 2 and 7029 DF,  p-value: < 2.2e-16
```

```
## multicollinearity
vif(mod1)
```

```
##       type enrollment
##   1.413965   1.413965
```

```
##We have not met the assumption of homoskedasticity because there is a signi
ficant difference in residuals at different levels of the predictors.

##The vif value for type of school and enrollment is approximately 1.5, which
indicates low correlation among variables. This model does not suffer from mu
lticollinearity.
```

QUESTION 4 [8 pts]

Recenter the variable enrollment at its mean. [2 pts] Create an interaction between type (type) and student enrollment (enrollment) recentered and rerun the model predicting exempt (exempt). [2 pts] Assume that type moderates the effect of enrollment in your interpretation of the interaction. (i) Interpret the results on each coefficient. (ii) Create a plot to visualize the interaction. [4 pts]

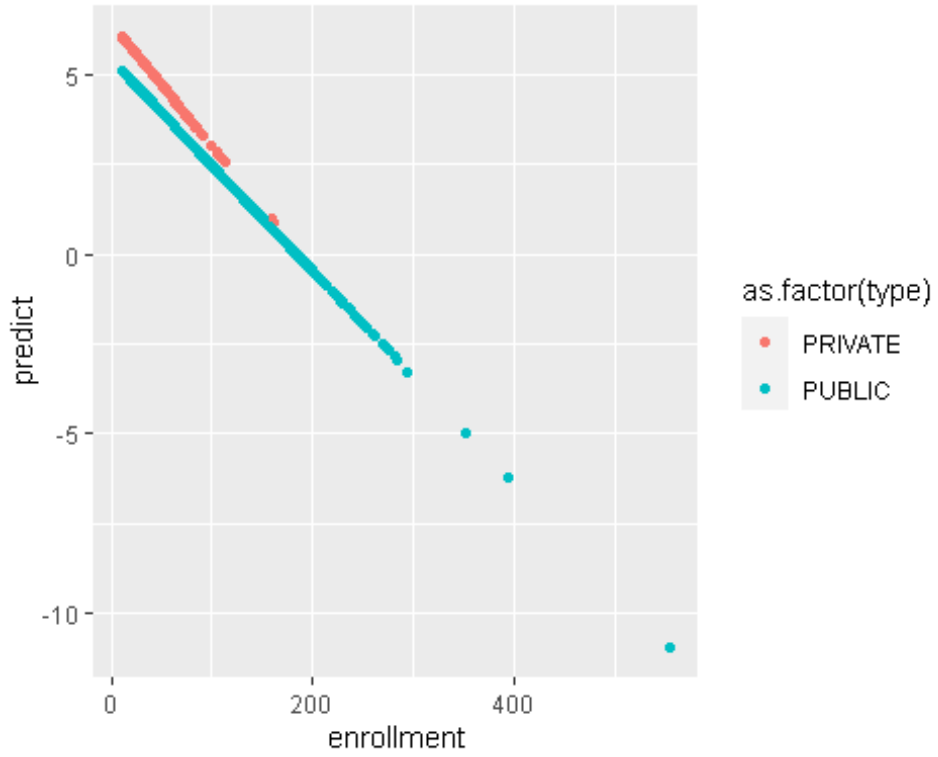
```
cavax$enrol_rec <- cavax$enrollment - mean(cavax$enrollment)
mean(cavax$enrollment)
```

```
## [1] 75.64562
```

```
mod3 <- lm(exempt ~ type * enrol_rec, data = cavax)
summary(mod3)
```

```
##
## Call:
## lm(formula = exempt ~ type * enrol_rec, data = cavax)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.070 -3.293 -1.726  0.587 86.926
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.837202   0.600435   6.391 1.76e-10 ***
## typePUBLIC           -0.651213   0.609131  -1.069  0.28507
## enrol_rec            -0.034017   0.011800  -2.883  0.00395 **
## typePUBLIC:enrol_rec  0.004599   0.012048   0.382  0.70272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.396 on 7028 degrees of freedom
## Multiple R-squared:  0.04036,    Adjusted R-squared:  0.03995
## F-statistic: 98.53 on 3 and 7028 DF,  p-value: < 2.2e-16
```

```
#There are a number of options to create an interaction plot.  Below, I will
show two.
#In addition to these two you can use the expand.grid() approach we used in c
lass as well.
cavax$predict = predict(mod3)
ggplot(cavax, aes(x = enrollment, y = predict, color = as.factor(type))) + ge
om_point()
```

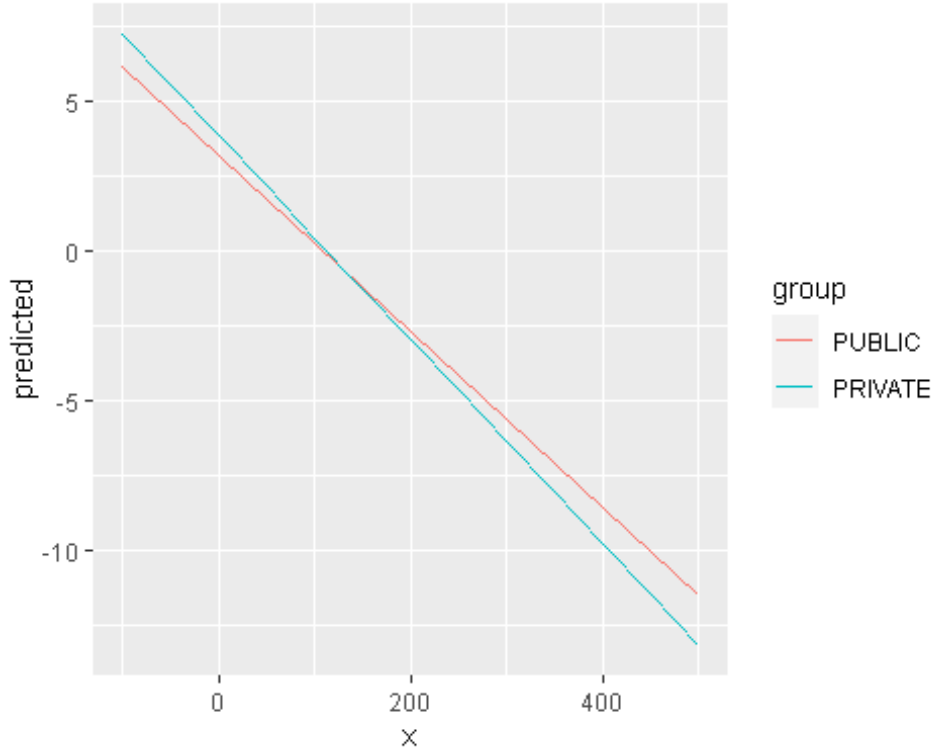



```
#can do this using the ggeffects package as well
dat = ggpredict(mod3, terms = c("enrol_rec", "type"))
dat
```

```
## # Predicted values of exempt
## # x = enrol_rec
##
## # type = PUBLIC
##
##    x | Predicted |          95% CI
## ----------------------------------
## -100 |      6.13 | [  5.56,  6.69]
##    0 |      3.19 | [  2.98,  3.39]
##  100 |      0.24 | [ -0.22,  0.71]
##  200 |     -2.70 | [ -3.62, -1.78]
##  300 |     -5.64 | [ -7.03, -4.25]
##  500 |    -11.52 | [-13.86, -9.19]
##
## # type = PRIVATE
##
##    x | Predicted |          95% CI
## ----------------------------------
## -100 |      7.24 | [  5.98,  8.50]
##    0 |      3.84 | [  2.66,  5.01]
##  100 |      0.44 | [ -3.01,  3.88]
##  200 |     -2.97 | [ -8.72,  2.78]
```

```
##  300 |     -6.37 | [-14.43,  1.69]
##  500 |    -13.17 | [-25.85, -0.49]
```

```
ggplot(dat, aes(x = x, y = predicted, color = group)) + geom_line()
```




```
##The base group is private school with 76 students enrolled in kindergarten.
##The coefficient for type is -0.65. Because type is part of an interaction,
we have to interpret it as simple main effect. Public schools with 76 student
s enrolled in kindergarten have 0.65 percentage points less exemptions than p
rivate schools with 76 students enrolled in kindergarten. The effect is not s
ignificant (p>.05)

##Enrollment is part of an interaction, so we have to interpret it as simple
main effect. The coefficient for enrollment shows the effect of enrollment on
ly on private schools. The coefficient is -0.03 which means that private scho
ols have 0.03 percentage points less students exemptions for each additional
student enrolled. The relationship is significant (p<.01).

##The coefficient for the interaction is 0.005. This value is the difference
in effect of enrollment between public and private schools. For public school
s the effect of enrollment is 0.005 percentage points more than the effect of
enrollment in private schools. This means that public schools have 0.005 perc
entage points more student exemptions for each additional student enrolled in
kindergarten compared to private schools.  The effect of enrollment in public
schools is -.034 + .0046 =  -.029
```

QUESTION 5 [10 pts]

Let's log transform (using the natural log) the variable enrollment and call the new variabe log_enroll. [2 pts] Estimate a model predicting exempt (exempt) by district type (type) and log of enrollment (log_enrollment). [2pts] Interpret the coefficient on the log of enrollment. [2 pts] Does it make more sense to use enrollment or the log of enrollment as the predictor variable? Why? [4 pts]

```
cavax$log_enroll = log(cavax$enrollment)

mod4 = lm(exempt ~ type + log_enroll, data = cavax)
summary(mod4)

##
## Call:
## lm(formula = exempt ~ type + log_enroll, data = cavax)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.417 -2.982 -1.461  0.867 85.934
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.1408     0.5410  26.137   <2e-16 ***
## typePUBLIC    0.5707     0.2878   1.983   0.0474 *
## log_enroll   -2.7338     0.1589 -17.201   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.324 on 7029 degrees of freedom
## Multiple R-squared:  0.05888,    Adjusted R-squared:  0.05861
## F-statistic: 219.9 on 2 and 7029 DF,  p-value: < 2.2e-16

##Δy = β1Δlog(x), we can simplify it with our approximation that Δy = (β1/100
)(%Δx). So, β1/100, which is -2.7338/100 = -0.027338. A 1% change in enrollme
nt causes exempt to decrease by 0.027 percentage points. The relationship is
significant (p<.001). Note, don't be confused by the fact that the dependent
variable in this model is a percentage from 0 to 100.  This is a lin-log mode
l, so the percentage change in the predictor variable as a unit change on the
dependent variable.  Here a unit change is now in terms of percentage points.
So a 1% change in enrollment decreases exempt by .027 percentage points.

#You can also calculate this effect "by hand".
#
#Let's calculate the exact percentage change
#what we need to do now is look at a 1% change in your level of
#enrollment.  So let's say we are currently at 500 students, a
#1% increase would bring us to 505 students
a = 14.14 + .5707 + -2.7338*log(500)
b = 14.14 + .5707 + -2.7338*log(500 + .01*500)
```

```
#Now in this model, we interpret the percentage change in the predictor
#on the unit change in crime.  Thus all we need to do is take the following:
b-a #and we get -.027; approximately B1/100 unit change.
```

```
## [1] -0.02720221
```

```
#Why might a logged predictor make sense? Logged predictors are useful becaus
e we can capture potential non-linear relationships.  For instance, you may n
ot think that an increase of 10 students would have the same impact on exempt
ions in all schools.  For smaller schools, 10 students is a big increase in e
nrollment.  For large schools it is not much of a change at all.  So it may b
e better to look at the percentage change in enrollment rather than the actua
l change in the number of students.  So for a school with 100 a 1 percent cha
nge is 1 student.  For a school with 1000 a 1 percent change is 10 students.
```

QUESTION 6 [12 pts]

Create a binary variable to indicate high versus low exempt rates. For schools with exempt percentages equal to or greater than 33 percent, indicate them as "high", for all other schools indicate them as "low". [2 pts] Run a logistic regression predicting whether a school is high versus low, in other words, we want our model to predict schools falling into the high category. In your model use the predictors of school type (type) and enrollment (enrollment) (note: do ***not*** *use log_enroll in this model). [2 pts] Interpret the coefficients on type and enrollment in terms of both log odds and odds. [6 pts] What is the probability of being a high exempt school if the school is private and has 100 students enrolled? [2 pts]*

```
#The easiest way to do this is to create a binary variable (0,1) and use the
value of
#'1' to indicate a high exempt school.
cavax$exempt_binary <- ifelse(cavax$exempt >= 33, 1, 0)

mod5 <- glm(factor(exempt_binary) ~ type + enrollment, family = binomial(link
= "logit"), data = cavax)
summary(mod5)
```

```
##
## Call:
## glm(formula = factor(exempt_binary) ~ type + enrollment, family = binomial
(link = "logit"),
##     data = cavax)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3496  -0.2117  -0.1293  -0.0801   4.5249
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.569319   0.174010 -14.765  < 2e-16 ***
## typePUBLIC   0.115309   0.232708   0.496     0.62
## enrollment  -0.031009   0.004197  -7.388 1.49e-13 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1108.0  on 7031  degrees of freedom
## Residual deviance: 1005.3  on 7029  degrees of freedom
## AIC: 1011.3
##
## Number of Fisher Scoring iterations: 8
```

```
exp(mod5$coef)
```

```
## (Intercept)  typePUBLIC  enrollment
##  0.07658769  1.12222035  0.96946696
```

```
##The coefficient for public schools is 0.115309 which means the log odds of being a high exempt school is 0.115309 greater for public schools compared to private schools. The exponentiated coefficient on public schools is 1.1; whic h means the odds of being a high exempt school are  by 1.12 times greater for public schools compared to private schools. The relationship is not significa nt (p>.05)

##The coefficient for enrollment is -0.03, which means the log odds of being a high exempt school decreases by 0.03 for each additional student enrolled i n kindergarten. The exponentiated coefficient of enrollment is 0.97 which mea ns the odds of being a high exempt school decrease by 3% for each additional student enrolled in kindergarten. The relationship is significant (p<.001).


##probability of being a high exempt school if the school is public and has 1 0 students enrolled
y3 = -2.569319 + .115 + (-0.031009*10)
y3
```

```
## [1] -2.764409
```

```
exp(y3)/(1+exp(y3))
```

```
## [1] 0.05927802
```

```
##a different way
s1=as.matrix(c(1,1,10))
odds_s1=exp(crossprod(s1, coefficients(mod5)))
odds_s1/(1+odds_s1)
```

```
##            [,1]
## [1,] 0.05929534
```

```
##The probability of being high exempt school if the schools is public and ha s 10 students is .059, or 5.9%.
```

PART TWO

Read in the 'concealed_carry' data. The data can be used to explore the relationship between adoption of concealed carry laws that allow to carry a concealed firearm. The data contain the following variables:

Variable name (Description)

- *stateid* (State id)
- *statename* (Name of state)
- *shall* (Equals 1 if state has concealed carry on that year and 0 if not)
- *year* (Year)
- *vio* (Violent crime rate per 100,000 people)
- *mur* (Murder rate per 100,000 people)

QUESTION 7 [10 pts]

a. Let's begin by exploring the data. How many years are there in the concealed_carry data? How many observations per state? [2 pts]

```
table(concealed_carry$year)

##
## 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991
1992
##   51   51   51   51   51   51   51   51   51   51   51   51   51   51   51
51
## 1993 1994 1995 1996 1997 1998 1999
##   51   51   51   51   51   51   51

table(concealed_carry$statename)

##
##              Alabama               Alaska              Arizona
##                   23                   23                   23
##             Arkansas           California             Colorado
##                   23                   23                   23
##          Connecticut             Delaware District of Columbia
##                   23                   23                   23
##              Florida              Georgia               Hawaii
##                   23                   23                   23
##                Idaho             Illinois              Indiana
##                   23                   23                   23
##                 Iowa               Kansas             Kentucky
##                   23                   23                   23
##            Louisiana                Maine             Maryland
##                   23                   23                   23
##        Massachusetts             Michigan            Minnesota
##                   23                   23                   23
##          Mississippi             Missouri              Montana
```

```
##                  23                  23                  23
##            Nebraska              Nevada       New Hampshire
##                  23                  23                  23
##          New Jersey          New Mexico      New York     NY
##                  23                  23                  23
##      North Carolina        North Dakota                Ohio
##                  23                  23                  23
##            Oklahoma              Oregon        Pennsylvania
##                  23                  23                  23
##        Rhode Island      South Carolina        South Dakota
##                  23                  23                  23
##           Tennessee               Texas                Utah
##                  23                  23                  23
##             Vermont            Virginia          Washington
##                  23                  23                  23
##       West Virginia           Wisconsin             Wyoming
##                  23                  23                  23
```

b. How many states had concealed carry laws in 1977 and how many had concealed carry laws in 1999? [4pts]

```
concealed_carry1 = concealed_carry %>%
  filter(year == 1977) %>%
  count(shall)
concealed_carry1
```

```
## # A tibble: 2 x 2
##   shall     n
##   <dbl> <int>
## 1     0    47
## 2     1     4
```

```
concealed_carry2 = concealed_carry %>%
  filter(year == 1999) %>%
  count(shall)
concealed_carry2
```

```
## # A tibble: 2 x 2
##   shall     n
##   <dbl> <int>
## 1     0    22
## 2     1    29
```

c. Create a plot tracking the violent crime rate over time for states that have ever adopted conceal carry laws and those that have never adopted the law. [4 pts]

```
concealed_carry4 = concealed_carry %>%
  group_by(statename) %>%
  summarise(years_shall = sum(shall))
```

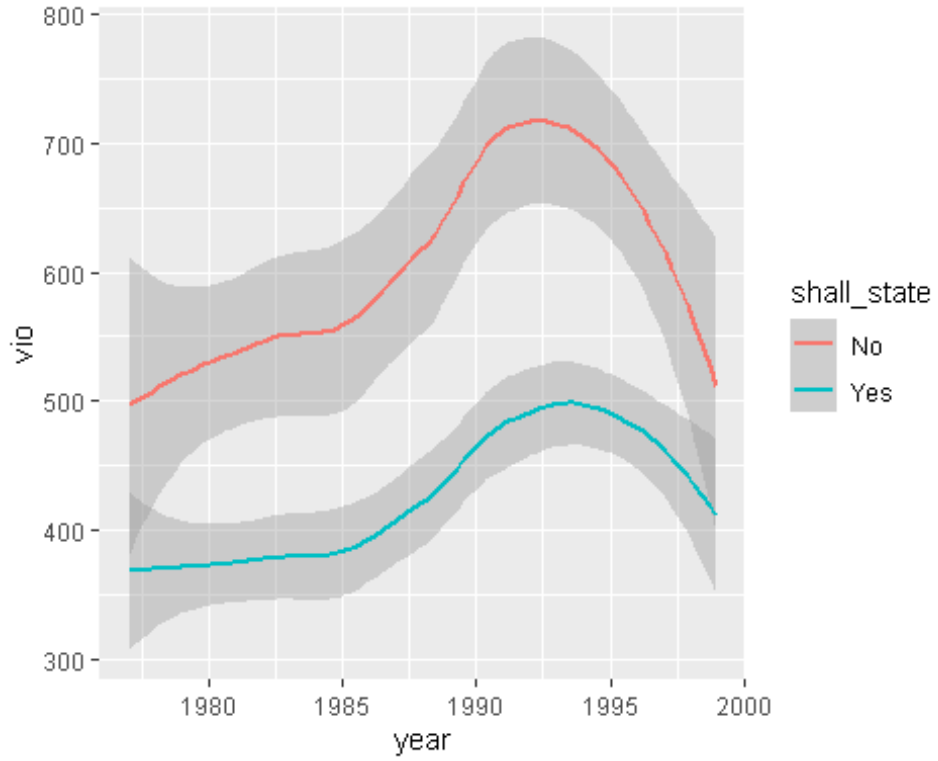
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
concealed_carry4$shall_state <- ifelse(concealed_carry4$years_shall > 0, "Yes
", "No")


concealed_carry5 = left_join(concealed_carry, concealed_carry4, by="statename
")

concealed_carry5$shall_state <- as.factor(concealed_carry5$shall_state)
ggplot(concealed_carry5, aes(x=year, y =vio, color = shall_state))+ geom_smoo
th()

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



QUESTION 8 [10 pts]

Convert the violent crime rate (vio) into a logged variable (using the natural log), call it log_vio. [2 pts] This will be our dependent variable. Run a pooled regression of the data (i.e., standard ols model as if this was cross-sectionall data) predicting the log of violent crimes (log_vio) as a function of the presence of concealed carry laws (shall) and a set of dummy variables for year. [2 pts] Iterpret the effect of shall. [2 pts] In general terms, what do the year dummy variables tell us about crime trends? [2 pts] In our current specificaiton of the model, is the effect of shall the same for all years? Why or why not? [2 pts]

```r
concealed_carry$log_vio = log(concealed_carry$vio)
mod7 <- lm(log_vio ~ shall + factor(year), data =concealed_carry)
summary(mod7)
```

```
##
## Call:
## lm(formula = log_vio ~ shall + factor(year), data = concealed_carry)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14538 -0.43698  0.05355  0.40556  1.68492
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.86727    0.08372  70.086  < 2e-16 ***
## shall             -0.59812    0.04446 -13.452  < 2e-16 ***
## factor(year)1978   0.04838    0.11829   0.409 0.682589
## factor(year)1979   0.14372    0.11829   1.215 0.224631
## factor(year)1980   0.18781    0.11829   1.588 0.112634
## factor(year)1981   0.17746    0.11829   1.500 0.133838
## factor(year)1982   0.14767    0.11829   1.248 0.212166
## factor(year)1983   0.08999    0.11829   0.761 0.446945
## factor(year)1984   0.10077    0.11829   0.852 0.394470
## factor(year)1985   0.12826    0.11829   1.084 0.278473
## factor(year)1986   0.20558    0.11832   1.738 0.082559 .
## factor(year)1987   0.19364    0.11834   1.636 0.102056
## factor(year)1988   0.24561    0.11837   2.075 0.038211 *
## factor(year)1989   0.28168    0.11837   2.380 0.017490 *
## factor(year)1990   0.41694    0.11849   3.519 0.000451 ***
## factor(year)1991   0.48647    0.11868   4.099 4.44e-05 ***
## factor(year)1992   0.52834    0.11883   4.446 9.59e-06 ***
## factor(year)1993   0.53923    0.11883   4.538 6.28e-06 ***
## factor(year)1994   0.51493    0.11883   4.333 1.60e-05 ***
## factor(year)1995   0.54706    0.11921   4.589 4.94e-06 ***
## factor(year)1996   0.54408    0.11983   4.540 6.21e-06 ***
## factor(year)1997   0.55553    0.12028   4.619 4.30e-06 ***
## factor(year)1998   0.49975    0.12028   4.155 3.50e-05 ***
## factor(year)1999   0.44010    0.12028   3.659 0.000265 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5973 on 1149 degrees of freedom
## Multiple R-squared:  0.161,  Adjusted R-squared:  0.1442
## F-statistic: 9.585 on 23 and 1149 DF,  p-value: < 2.2e-16
```

```
100*(exp(-0.6) - 1)
```

```
## [1] -45.11884
```

```
##The natural log of violent crime rate per 100,000 people decreases by .59 f
or the states that have concealed carry laws. The relationship is significant
(p<.001). This relationship is difficult to understand, so we will return to
thinking about semi-elasticities and percentage changes. The violent crime ra
te per 100,000 people is 45% lower when there are concealed carry laws.
```

```
#Overtime, we can see the violent crime  rate increase, especially through th
e mid 90s, and start to drop down again in the late 90s.

#The effect of shall is the same for all years.  Why - this is how we modeled
it.  If we wanted to see if shall had different effects in different years, w
e would need to create an interaction term.
```