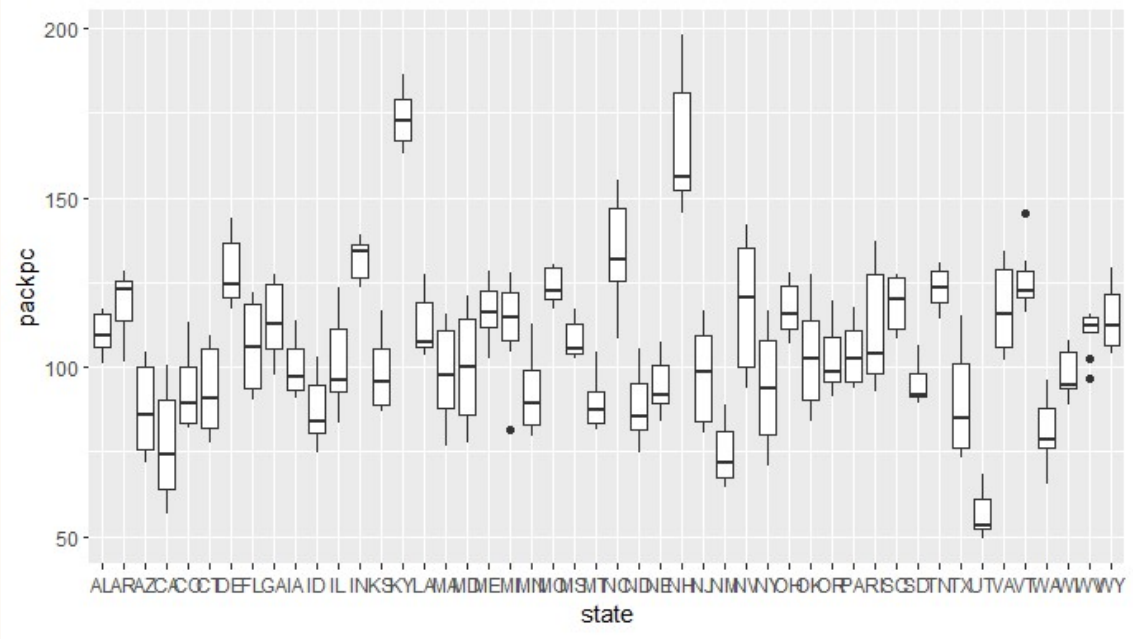
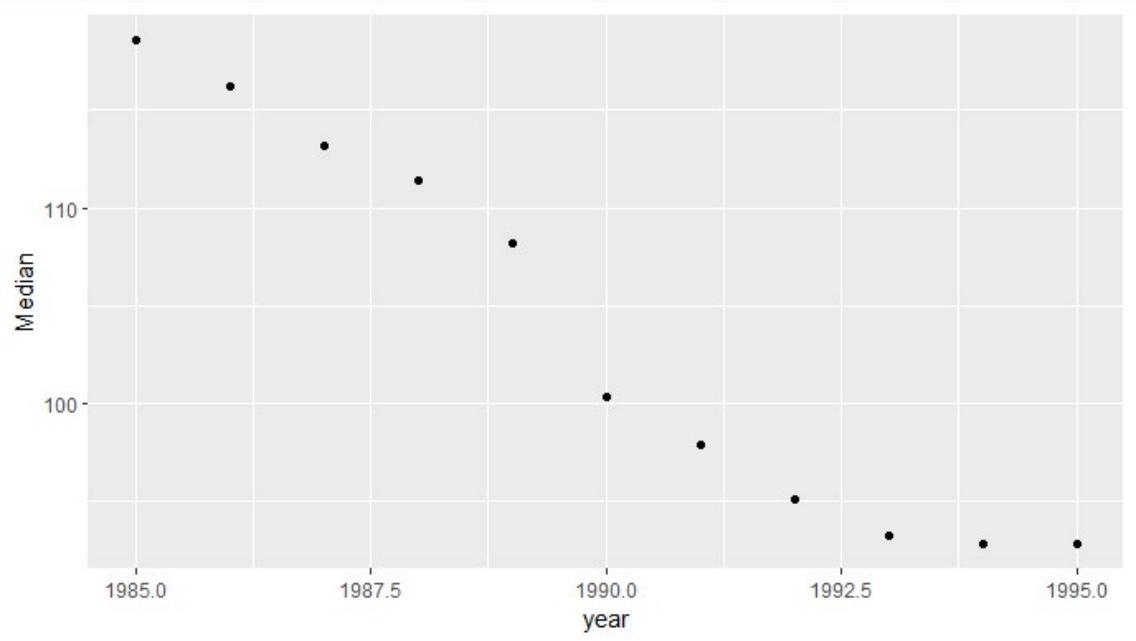
R Script Code:

```
# loading in libraries
install.packages("Ecdat")
library(Ecdat)
head(Cigarette)

library("ggplot2")
library("dplyr")

# Create a boxplot of the average number of packs per capita by state
ggplot(Cigarette, aes(x=state, y=packpc)) + geom_boxplot()

# Which state has the highest number of packs, and which state has the lowest
Cigarette %>% group_by(state) %>% summarise(Mean = mean(packpc)) %>% arrange(desc(Mean))
# Kentucky has the highest average number of packs per capita by state.
# Utah has the lowest average number of packs per capita by state.
```

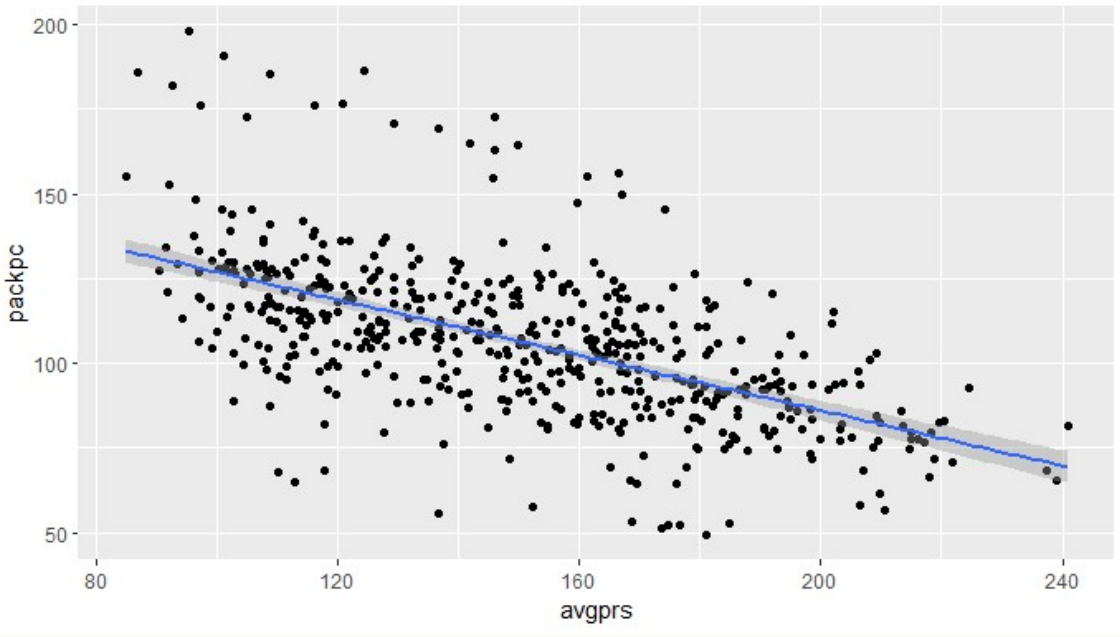



R Script Code:

```
# What is the median over all states of the number of packs per capita for each year?
MedianDF <- Cigarette %>% group_by(year) %>% summarise(Median = median(packpc))
# Plot the median value for the years 1985 to 1995.
unique(Cigarette$year)
ggplot(MedianDF, aes(x=year, y=Median)) + geom_point()
```



The Median over all states of the number of packs per capita for each declined from 1985 to 1995. This indicates that cigarette usage dropped from 1985 to 1995.



R Script Code:

```
# Create a scatter plot of price per pack versus the number of packs per capita for all states and years.
ggplot(Cigarette, aes(x=avgprs, y=packpc)) + geom_point() + geom_smooth(method=lm)

#Are the price and the per capita packs positively correlated, negatively correlated or uncorrelated?
cor.test(Cigarette$avgprs, Cigarette$packpc, method="pearson", use="complete.obs")
#Price and the per capita packs are moderately, negatively correlated.
```

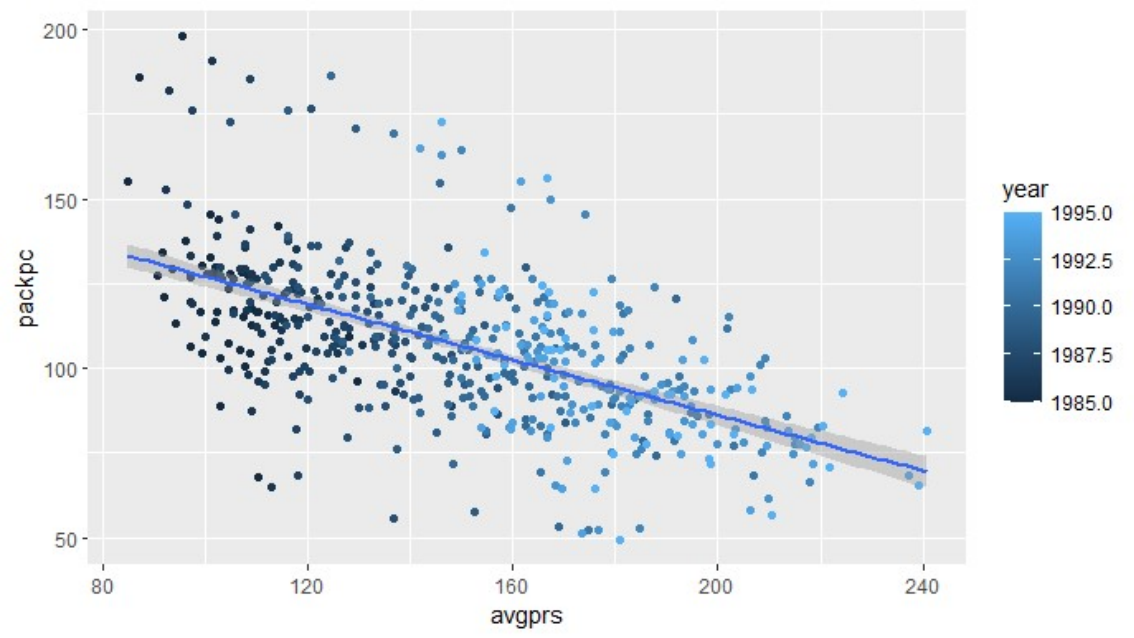
Console Output:

Pearson's product-moment correlation

data: Cigarette$avgprs and Cigarette$packpc
t = -16.562, df = 526, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.6388606 -0.5264104
sample estimates: cor
-0.5854443

Are the price and the per capita packs positively correlated, negatively correlated or uncorrelated?

The price and the per capita packs are moderately, negatively correlated. As price increases, per capita packs decreases.



R Script Code:

#Change the scatter plot to show the points for each year in a different color.

```
ggplot(Cigarette, aes(x=avgprs, y=packpc, color=year)) + geom_point() +
geom_smooth(method=lm)
```

#When the average price was lower in years 1985 to 1988, the per capita packs amount was higher. As the years progressed, the per capita packs amount continually decreased.

R Script Code:

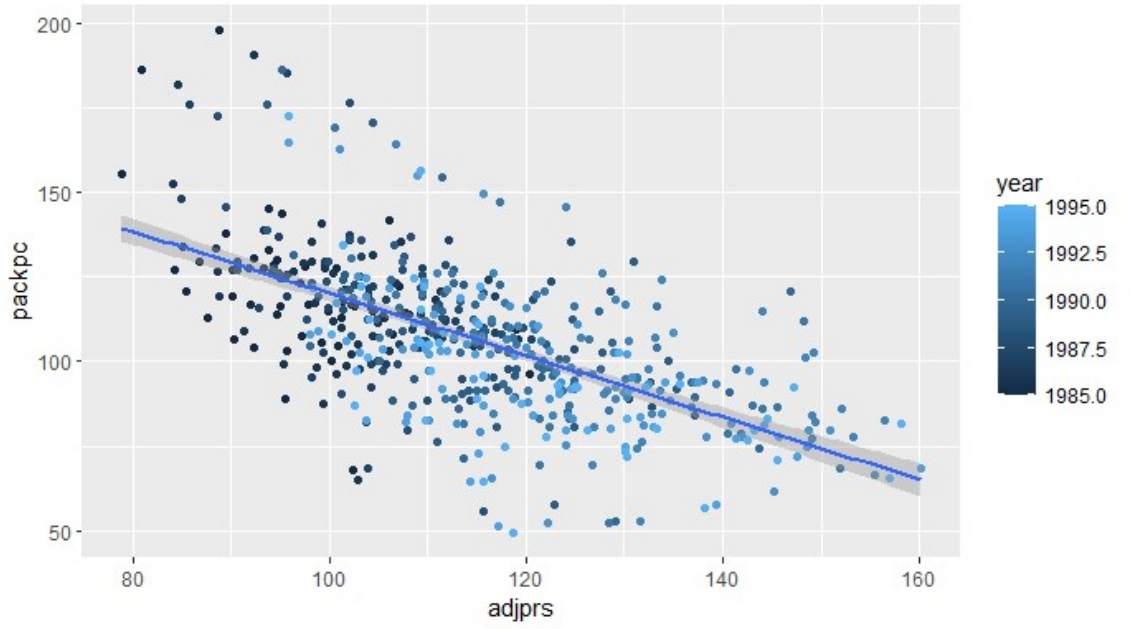
```
#Do a linear regression for these two variables.

regression <- lm(avgprs~packpc, Cigarette)
summary(regression)

#How much variability does the line explain?
#34% of the variability
```

Console Output:

```
Call:
lm(formula = avgprs ~ packpc, data = Cigarette)
Residuals:
     Min       1Q   Median       3Q      Max
-72.346  -20.729   -0.002   19.775   69.580
Coefficients:
             Estimate   Std. Error  t value  Pr(>|t|)
(Intercept)  239.50792   5.51461    43.43    <2e-16 ***
packpc       -0.83843   0.05062    -16.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.87 on 526 degrees of freedom
Multiple R-squared: 0.3427,     Adjusted R-squared: 0.3415
F-statistic: 274.3 on 1 and 526 DF,  p-value: < 2.2e-16
```

R Script Code:

```
#Adjust for inflation by dividing avgprs by cpi.
View(Cigarette)

AdjPrice <- Cigarette %>% mutate(adjprs =  avgprs / cpi)

#Create a scatter plot of Adjusted Price versus the number of packs per capita for all states and years.

ggplot(AdjPrice, aes(x=adjprs, y=packpc, color=year)) + geom_point() + geom_smooth(method=lm)

#Do a linear regression for Adjusted Price and packs per capita.
Adjregression <- lm(adjprs~packpc, AdjPrice)
summary(Adjregression)
```



R Script Code:

```
#Do a linear regression for Adjusted Price and packs per capita.
Adjregression <- lm(adjprs~packpc, AdjPrice)
summary(Adjregression)
```

#How much variability does the line explain?
#38% of the variability.
#Adjusting for inflation increased the amount of variability the line explains.

Console Output:

```
Call:
 lm(formula = adjprs ~ packpc, data = AdjPrice)

Residuals:
     Min       1Q   Median       3Q      Max
-29.089   -8.497   -0.437    7.232   37.708

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) 158.69970   2.51169    63.19    <2e-16 ***
packpc       -0.41124   0.02306   -17.84    <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.24 on 526 degrees of freedom
Multiple R-squared: 0.3769,     Adjusted R-squared: 0.3757
F-statistic: 318.1 on 1 and 526 DF,     p-value: < 2.2e-16
```

R Script Code:

```
# Create a data frame with just the rows from 1985.

Rows1985 <- Cigarette %>% filter(year == 1985)

# Create a data frame with just the rows from 1995

Rows1995 <- Cigarette %>% filter(year == 1995)

#Get a vector of the number of packs per capita from each data frame
Packs1985 <- Rows1985$packpc
Packs1995 <- Rows1995$packpc

# Use paired t-Test
#Two Samples
#H0 = number of packs per capita 1985 == number of packs per capita 1995
#H1 = number of packs per capita 1985 != number of packs per capita 1995
t.obj <- t.test(Packs1985, Packs1995, paired=TRUE)
t.obj
```

The p-value is less than 0.05, which means there is significant difference between the number of packs per capita in 1985 and the number of packs per capita in 1995.

Paired t-test

```
data: Packs1985 and Packs1995
t = 14.789,  df = 47,  p-value < 2.2e-16
alternative hypothesis: true mean
difference is not equal to 0
95 percent confidence interval:
      22.21151   29.20576
sample estimates:
mean difference
      25.70863
```

WHAT QUESTIONS COULD THIS DATA SET ANSWER?

TAX VS. PACK PER CAPITA

- Is there a relationship between average state, federal and local taxes, and average number of packs of cigarettes per capita per year?

TAX VS. AVERAGE PRICE

- Is there a relationship between average state, federal and local taxes, and average price per pack of cigarettes?

INCOME VS. PACK PER CAPITA

- Is there a relationship between Personal Income and the average number orf packs of cigarettes per capita per year?

R Script Code:

```
# Create a scatter plot of tax versus the number of packs per capita for all states and years.
ggplot(Cigarette, aes(x=tax, y=packpc)) + geom_point() + geom_smooth(method=lm)

#Are the tax and the per capita packs positively correlated, negatively correlated or uncorrelated?
cor.test(Cigarette$tax, Cigarette$packpc, method="pearson", use="complete.obs")
```

Pearson's product-moment correlation

data: Cigarette$tax and Cigarette$packpc
t = -15.817, df = 526, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: -0.6228942 -0.5069682
sample estimates:
cor
-0.5677393

The tax and the per capita packs are moderately, negatively correlated. As tax increases, per capita packs decreases.