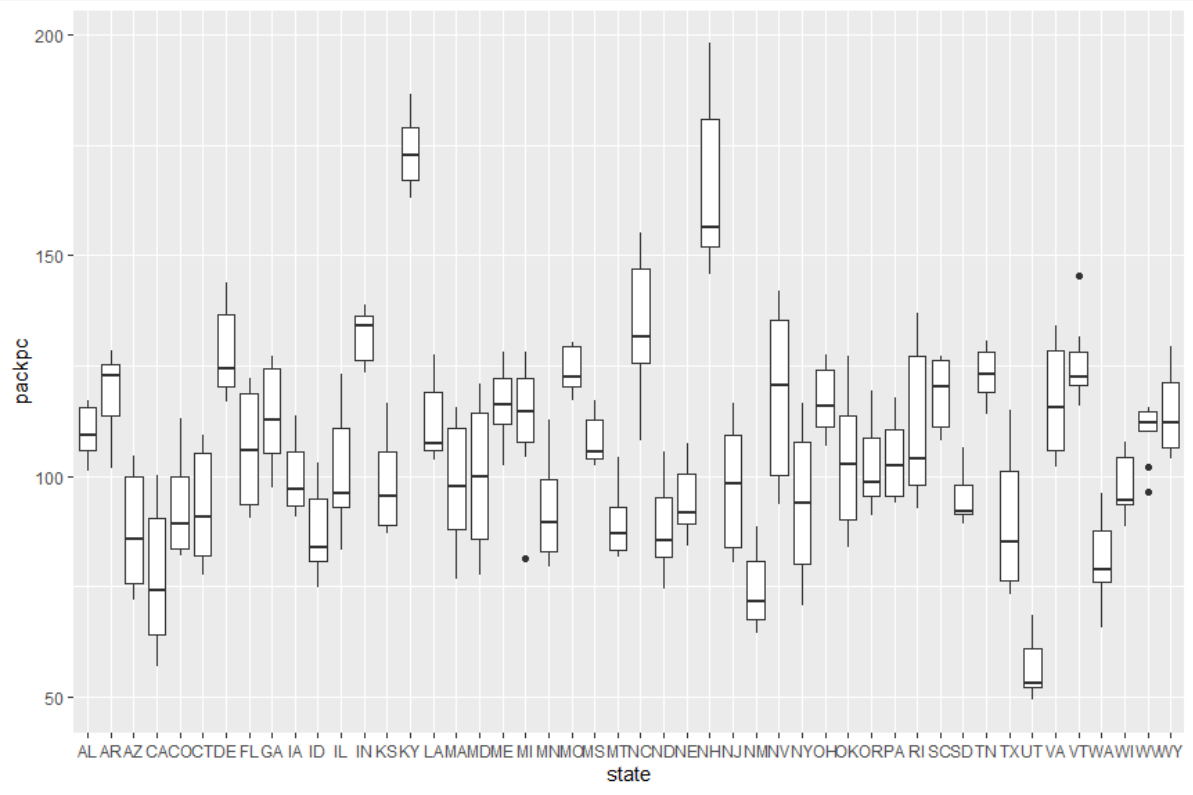
Lesson 10: Final Project

RICHMOND ANKU

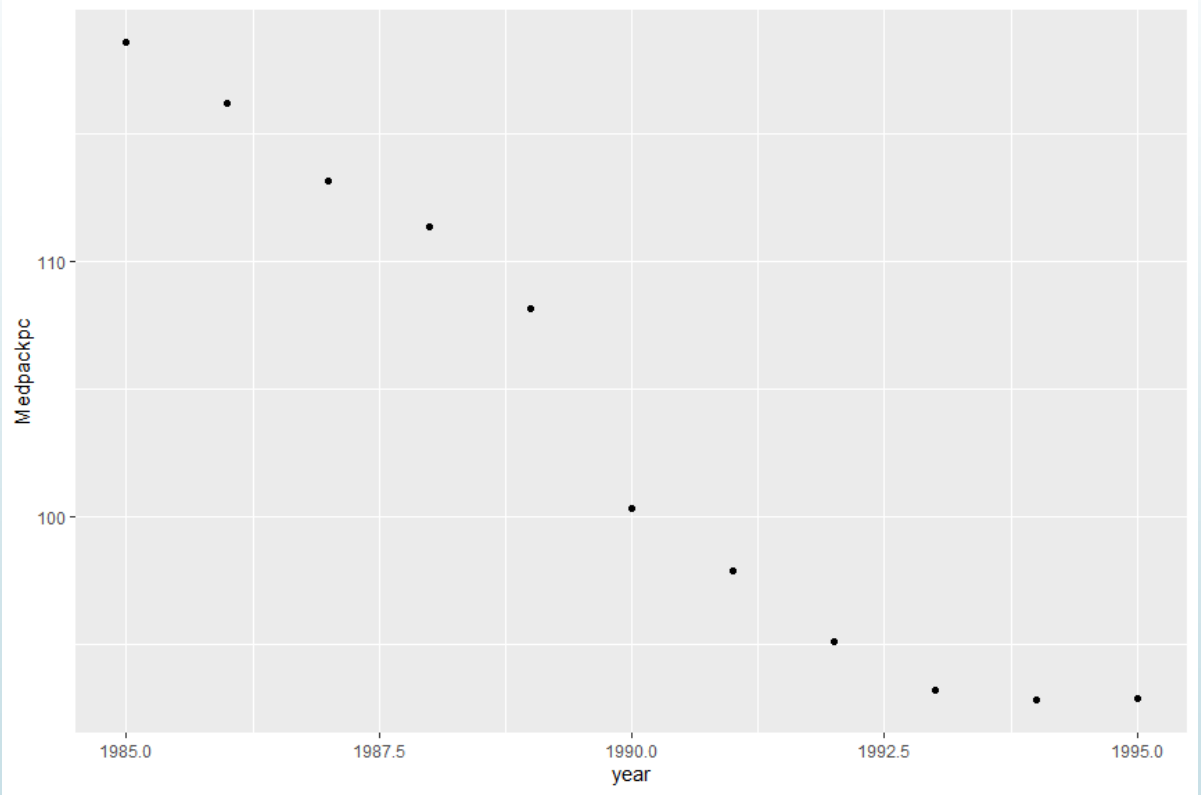
Boxplot of average number of packs per capita by State

- ggplot(Cigarette, aes(x= state, y= packpc)) + geom_boxplot()
- Extremes <- Cigarette %>% group_by(state) %>% summarise(Mean = mean(packpc)) %>% arrange(Mean)
- **UT** has the lowest number of packs per capita
- **KY** has the highest number of packs per capita



Median for number of packs per capita for all states by year

- ggplot(packpc_median, aes(x= year, y= Medpackpc)) + geom_point()
- From the graph, cigarette usage has gradually decreased over the years

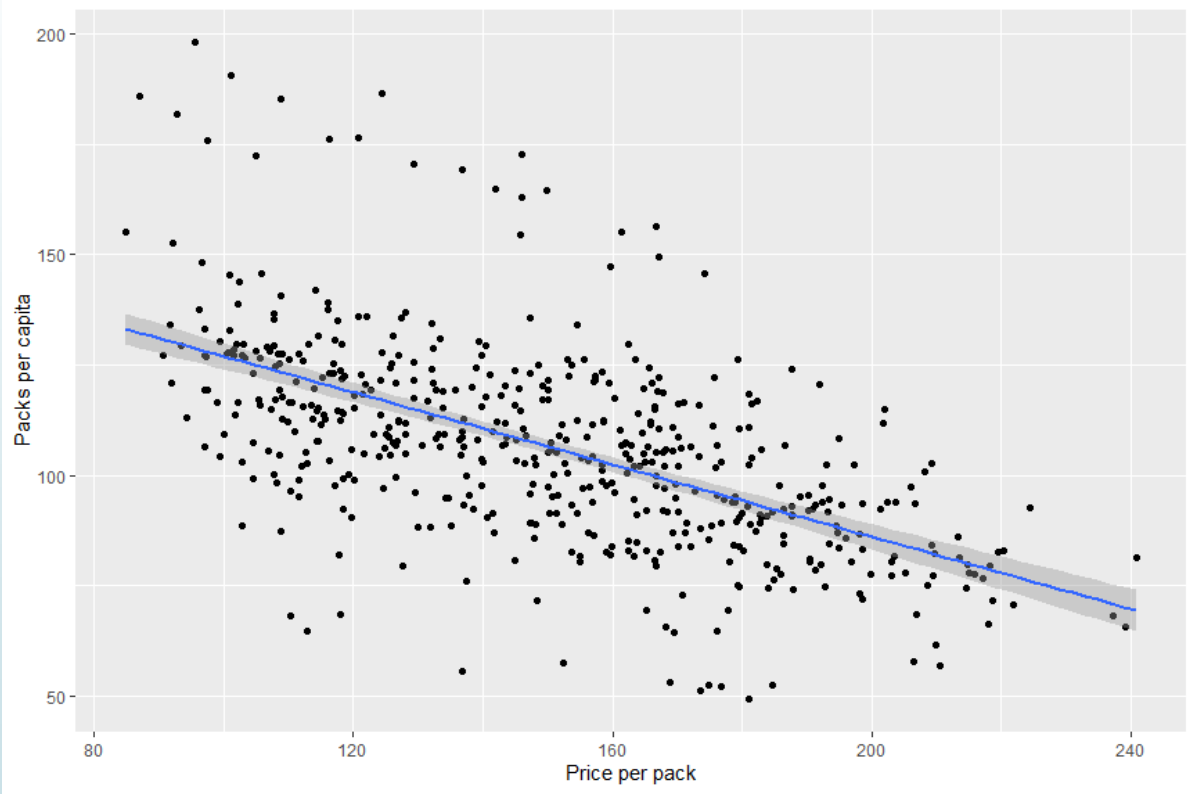


Scatter plot of price per pack vs number of packs per capita for all states and years

- ggplot(Cigarette, aes(x = avgprs, y = packpc)) + geom_point() + geom_smooth(method = lm) + xlab("Price per pack") + ylab("Packs per capita")

- cor.test(Cigarette$avgprs, Cigarette$packpc, method = "pearson", use = "complete.obs")

- There is **negative moderate** correlation between price & packs per capita



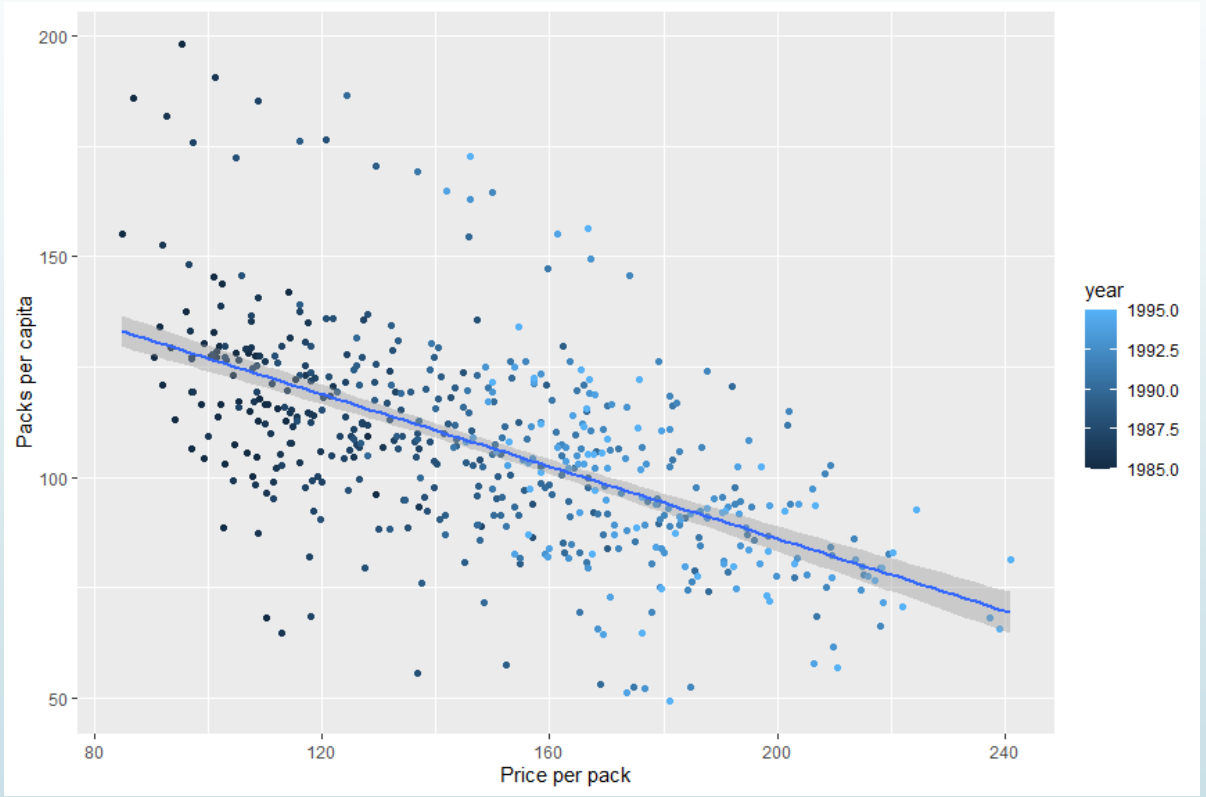
Scatter plot showing points for each year by color

- ggplot(Cigarette, aes(x = avgprs, y = packpc, color = year)) + geom_point() + geom_smooth(method = lm) + xlab("Price per pack") + ylab("Packs per capita")

- The 1980s saw higher cigarette consumption at lower prices. However, into the 1990s, as prices increased packs per capita decreased

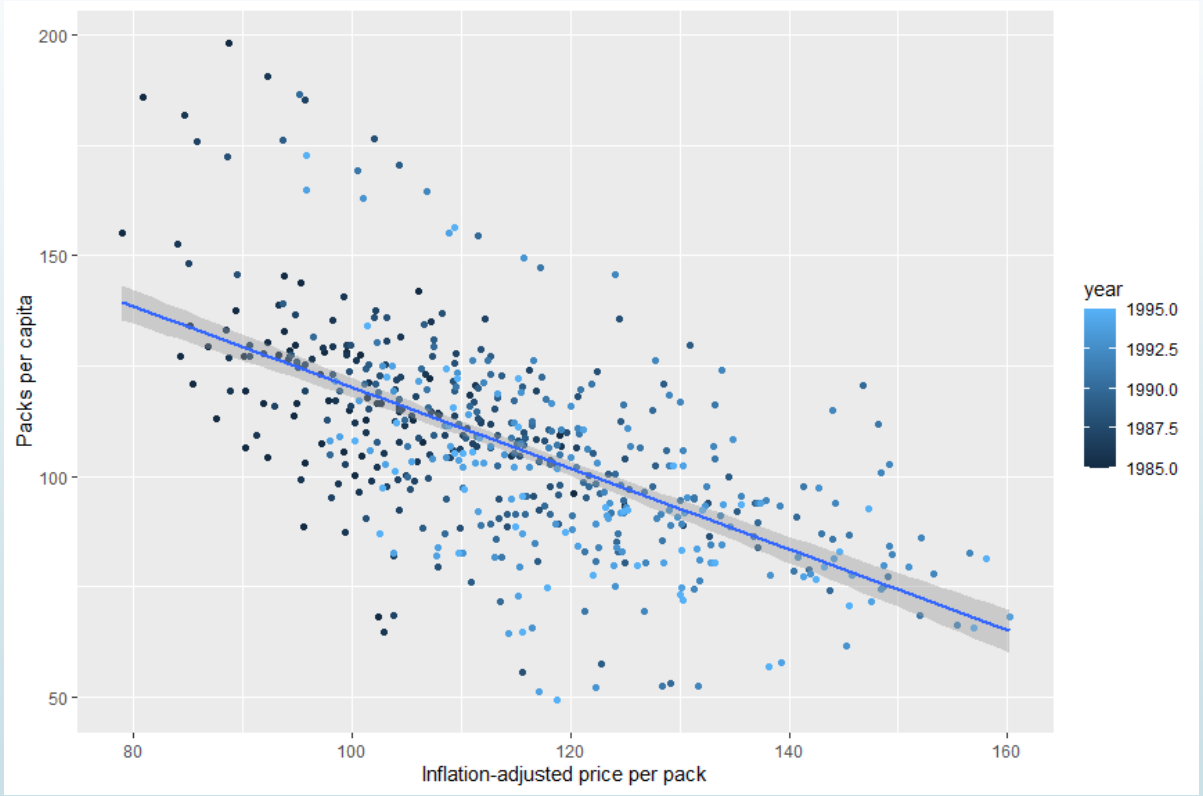
- regression <- lm(packpc~avgprs, Cigarette)
  summary(regression)

- The line explains **34%** of variability



Plotting scatter plot with inflation adjusted price & packs per capita

- infladjusted <- Cigarette %>% mutate(inflprs = avgprs / cpi)
- ggplot(infladjusted, aes(x = inflprs, y = packpc, color = year)) + geom_point() + geom_smooth(method = lm) + xlab("Inflation-adjusted price per pack") + ylab("Packs per capita")
- The line explains **34%** of variability



Using t-test to compare differences in packs per capita in 1985 &1995

- Cigarette1985 <- Cigarette %>% filter(year == 1985)
- Cigarette1995 <- Cigarette %>% filter(year == 1995)
- t.test(Cigarette1985$packpc, Cigarette1995$packpc, paired = TRUE)

- Mean1985 <- mean(Cigarette1985$packpc)
- The average packs per capita in 1985 is **122.04**
- Mean1995 <- mean(Cigarette1995$packpc)
- The average packs per capita in 1995 is **96.33**

- **#There was significant change in the number of packs per capita, decreasing from 122.04 packs per capita in 1985, to 96.33 in 1995**