Econometrics Hw2

Trevor Freeland

April 15, 2018

```
library(tidyverse)
library(pander)
library(googlesheets)
food.data <- read.csv("~/R/Econometrics/EconometricsHW2/FoodExpenditures.csv")
url <- gs_url("https://docs.google.com/spreadsheets/d/163zuhH_nUXtsDfnnrzuDxJOj4D3Ago1ldEA96OJZPmE/")
country.data <- gs_read(url)
```

1

(a)

T-stat = -1.77

P-value = .084

Conclusion: We cannot reject the null hypothesis.

Analysis: Assuming that our scores represent a sample from the population, we could not reject the null hypothesis that the true mean score for part 1 of the exam is 78%. Using a t distribution with 42 degrees of freedom we receive a p-value of .084 from a two-sided t-test. With a standard significance level of .05 we cannot reject the null hypothesis that the true mean is 78%.

```
null.mu <- .78
obs.mu <- 33.1/45
sd <- 7.4/45
n <- 43
t.stat <- (obs.mu - null.mu)/(sd/sqrt(n))
2 * pt(t.stat, df = 42)
```

```
[1] 0.08360196
```

(b)

99% confidence interval: Between -.356 and .50

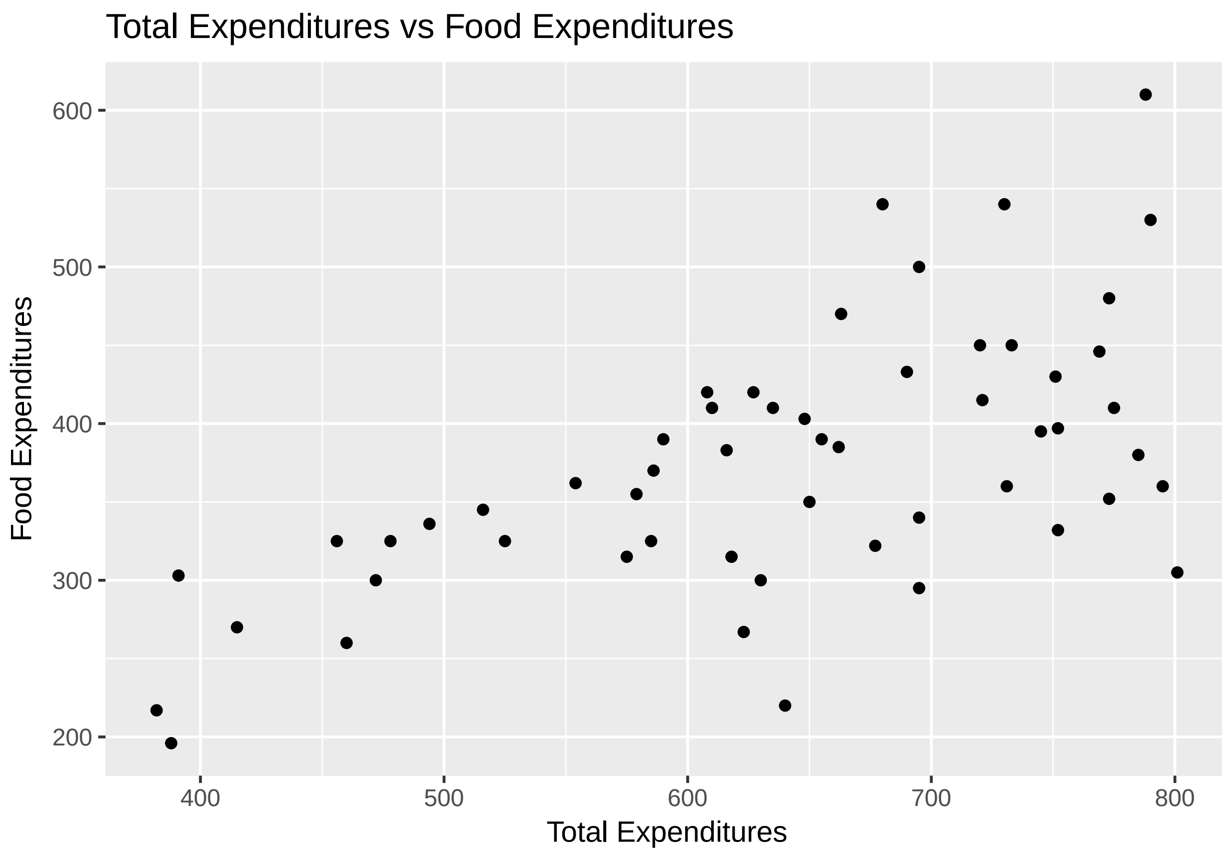
See code below for calculations.

```
coef <- .073
sd <- .159
upper <- coef + qt(.995, df=42)*sd
lower <- coef - qt(.995, df=42)*sd
```

2

(a)

```
ggplot(food.data) + geom_point(aes(x = TOTALEXP, y = FOODEXP)) + ggtitle("Total Expenditures vs Food Exp
```



(b)

Looking at the scatterplot above I am slightly worried about the linearity assumption we make in Classical Linear Regression. There appears to be some curvature in the data and so the relationship might not be best described with a linear model.

(c)

Our Model: Food Expenditures = 94.21 + .44(TotalExp) + ϵ (Rupees), ϵ~N(0,66.86^2)

Our intercept of 94.21 indicates that if someone had 0 Total Expenditures then they would still be expected to have 94.21 in food expenditures, which isn't all that helpful.

Our coefficient on Total Expenditures of .44 indicates that on average, for every 1 rupee increase in total expenditures, we would expect food expenditures to rise by .44 rupees.

See table below for model summary.

```
food.lm <- lm(FOODEXP~TOTALEXP, data = food.data)
pander(summary(food.lm))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	94.21	50.86	1.852	0.06953
TOTALEXP	0.4368	0.07832	5.577	8.451e-07

Table 2: Fitting linear model: FOODEXP ~ TOTALEXP

Observations	Residual Std. Error	R^2	Adjusted R^2
55	66.86	0.3698	0.3579

3

(a)

Our Model: Life Expectancy = 58.9 + .00069(gdpPercap) + ϵ, ϵ~N(0, 8.95^2)

Our intercept indicates that if somone was living in a place with 0 gdpPercap, on average they would be expected to live for 58.9 years.

Our coefficient for gdpPercap indicates that for an increase in gdpPercap of 1000 would lead to an average increase in life expectancy by .69 years.

See table below for model summary.

```
new.data <- country.data %>% filter(year %in% c(2000:2009))
country.lm <- lm(lifeExp~gdpPercap, data = new.data)
pander(summary(country.lm))
```

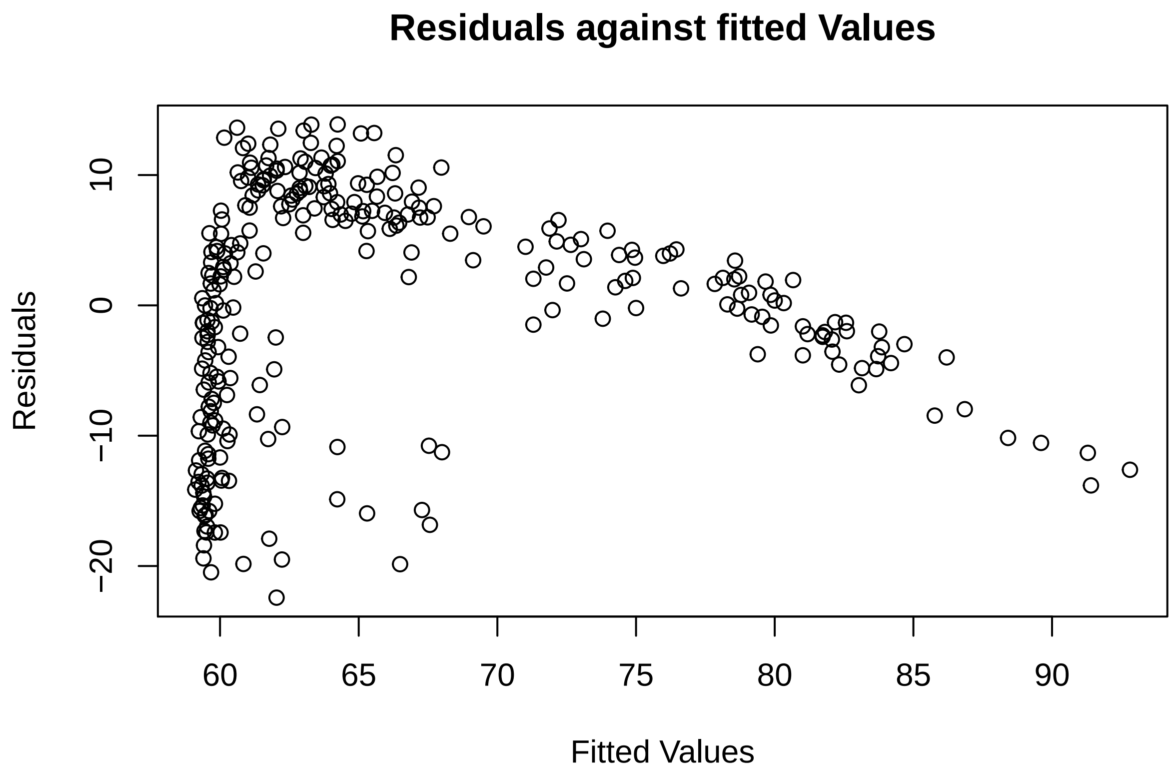
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	58.94	0.7138	82.57	1.382e-199
gdpPercap	0.0006862	4.416e-05	15.54	9.261e-40

Table 4: Fitting linear model: lifeExp ~ gdpPercap

Observations	Residual Std. Error	R^2	Adjusted R^2
284	8.951	0.4613	0.4594

(b)

```
plot(resid(country.lm)~fitted(country.lm), main = "Residuals against fitted Values", ylab="Residuals", :
```



I do have concerns about our model. There appears to be a very clear pattern that for larger fitted values, our residuals consistently get smaller and smaller, which means that our model is missing something and that some of our model assumptions might be wrong.

4

New Model: LifeExp = Intercept + ContinentIntercepts + gdpPercap + gdpPercap*Continent + ϵ, ϵ~N(0,σ^2)

With estimated Coefficients our mode looks like:

LifeExp = 50.7 + 19.3(Americas) + 14.2(Asia) + 21.3(Europe) + 25.1(Oceania) + .001(gdpPercap) + -.0009(Americas:gdpPercap) + -.0007(Asia:gdpPercap) + -.001(Europe:gdpPercap) + -.001(Oceania:gdpPercap) + ϵ, ϵ~N(0,6.3^2)

The difference in my model compared to our original model is that we have a different intercept for each continent, and each continent also has a different slope on gdpPercap. For example if we look at a country in the Americas, our expected LifeExp = 50.7 + 19.3 + (.001-.0009 = .0001)(gdpPercap).

My second model that I made fits the data better. I used the anova command, which compares two nested models, on the two models and my p-value for my anova test is essentailly 0, which indicates that we need to use the larger model, which is the model that I made.

See Table below for model summary and Anova results

```
country.lm2 <- lm(lifeExp~ continent * gdpPercap , data = new.data)
pander(summary(country.lm2))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	50.7	0.8196	61.86	5.084e-163
continentAmericas	19.35	1.565	12.36	3.407e-28
continentAsia	14.23	1.329	10.71	1.415e-22
continentEurope	21.33	2.031	10.5	6.584e-22
continentOceania	25.12	20.43	1.23	0.2199
gdpPercap	0.001183	0.0001885	6.276	1.349e-09
continentAmericas:gdpPercap	-0.0008909	0.0002122	-4.198	3.649e-05
continentAsia:gdpPercap	-0.000737	0.0001984	-3.715	0.000246
continentEurope:gdpPercap	-0.0009632	0.0002016	-4.778	2.886e-06
continentOceania:gdpPercap	-0.001028	0.0007352	-1.398	0.1633

Table 6: Fitting linear model: lifeExp ~ continent * gdpPercap

Observations	Residual Std. Error	R^2	Adjusted R^2
284	6.322	0.7389	0.7303

```
anova(country.lm2, country.lm)
```

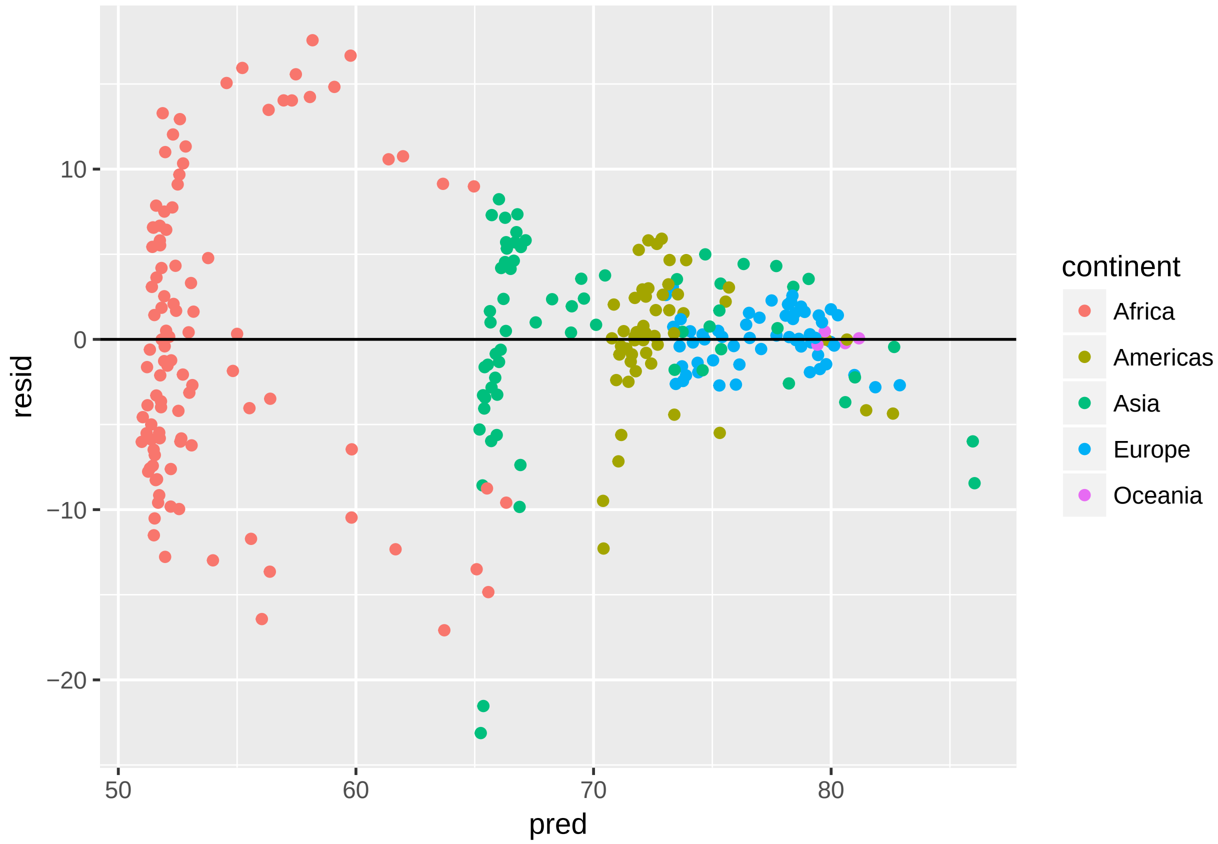
```
Analysis of Variance Table

Model 1: lifeExp ~ continent * gdpPercap
Model 2: lifeExp ~ gdpPercap
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    274 10950
2    282 22593 -8    -11643 36.419 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Challenge Problem

With the color we can see that our model does not do very well for countries in Africa. The model does a bit better for countries in Asia, but not extremely well still. The model then does a fairly good job at predicting for countries in Europe, the Americas, and Oceania.

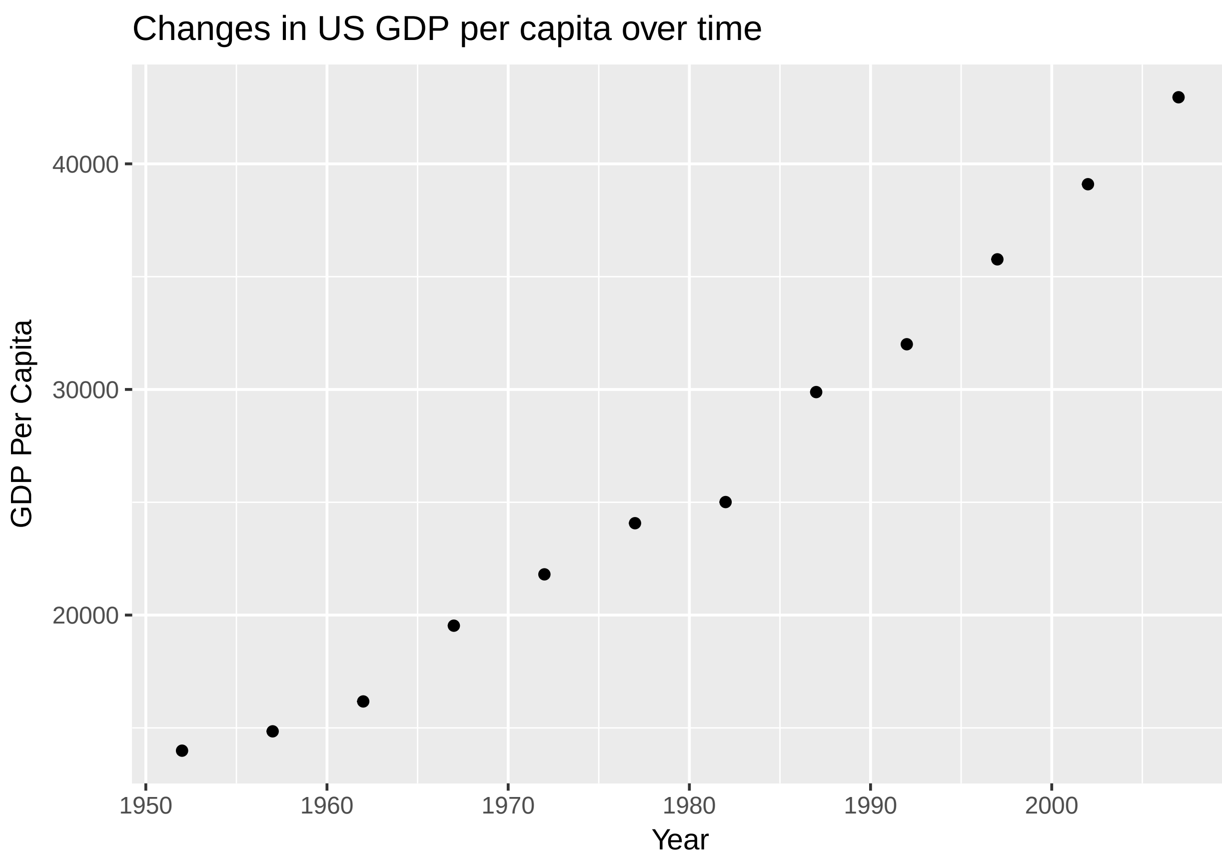
```
new.data$pred <- predict(country.lm2)
new.data$resid <- resid(country.lm2)
ggplot(new.data, aes(x = pred, y = resid, color = continent)) +  geom_point() + geom_abline(slope = 0, i
```



5

(a)

```
us.data <- country.data %>% filter(country == "United States")
ggplot(us.data) + geom_point(aes(x = year, y = gdpPercap)) + ggtitle("Changes in US GDP per capita over
```



(b)

On Average the GDP per capita in the US grew by \$533 per year.

```
us.lm <- lm(gdpPercap~year, data = us.data)
pander(summary(us.lm))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1027888	46934	-21.9	8.82e-10
year	532.5	23.71	22.46	6.882e-10

Table 8: Fitting linear model: gdpPercap ~ year

Observations	Residual Std. Error	R^2	Adjusted R^2
12	1418	0.9806	0.9786

(c)

A 95% confidence interval for the growth rate of GDP per capita per year in the US is between \$479.7 and \$585.4.

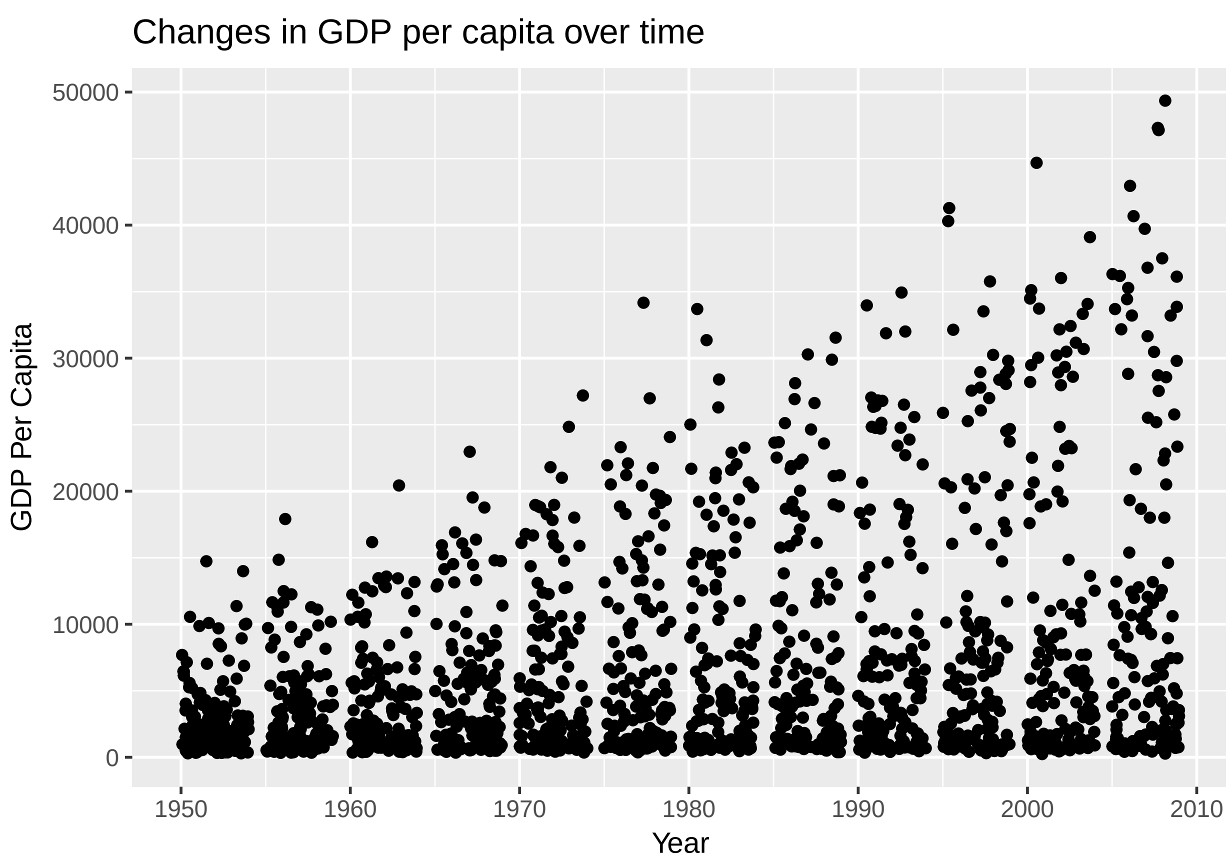
```r
confint(us.lm)
```

```
                    2.5 %       97.5 %
(Intercept) -1132463.9288 -923312.3500
year             479.7057     585.3605
```

6

(a)

```r
filtered.data <- country.data %>% filter(gdpPercap < 55000)
ggplot(filtered.data, aes(x = year, y = gdpPercap)) + geom_jitter() + ggtitle("Changes in GDP per capita
```



(b)

On Average the GDP per capita grows by $147.2 per year. See below for model summary and the regression line added to our plot from above.

```r
model1 <- lm(gdpPercap~year, data = filtered.data)
pander(summary(model1))
```

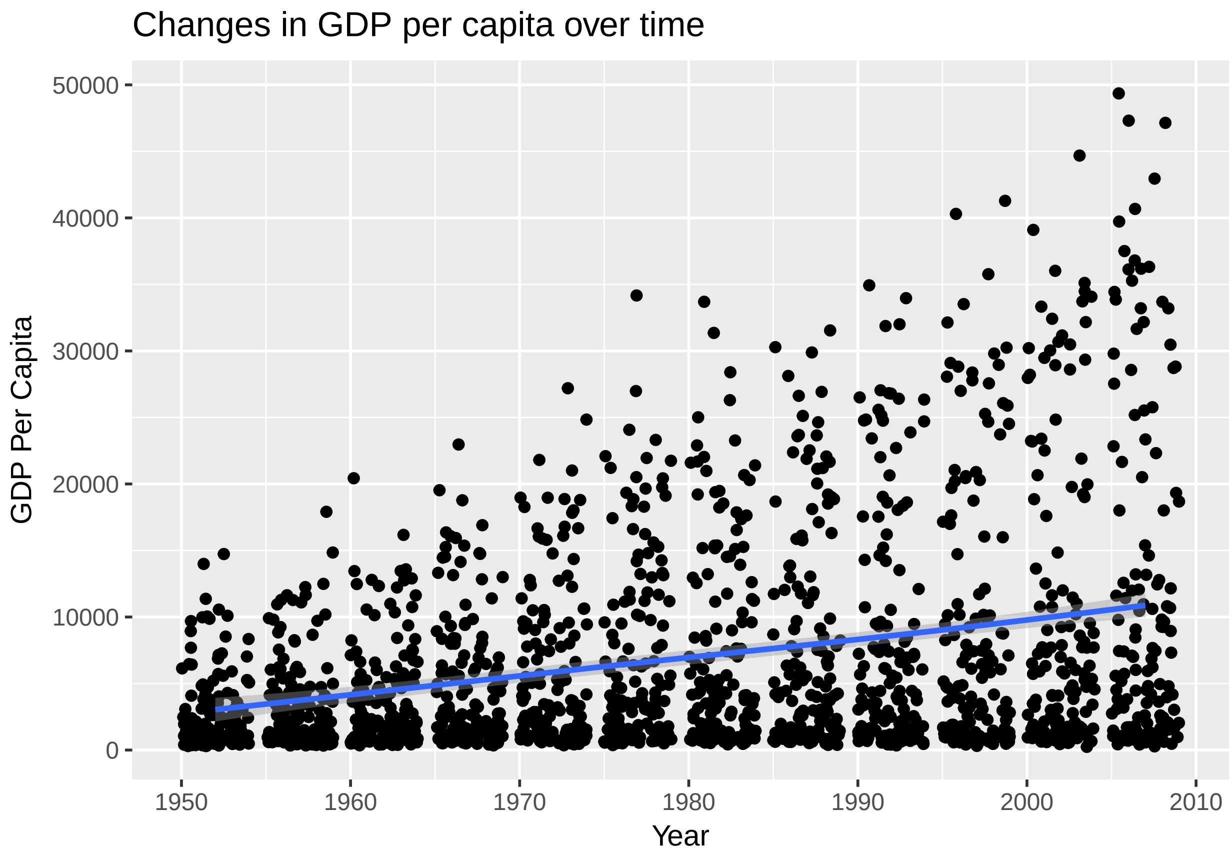
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-284429	22053	-12.9	2.245e-36
year	147.2	11.14	13.21	5.319e-38

	Estimate	Std. Error	t value	Pr(>|t|)

Table 10: Fitting linear model: gdpPercap ~ year

Observations	Residual Std. Error	R^2	Adjusted R^2
1698	7923	0.09331	0.09277

```
ggplot(filtered.data, aes(x = year, y = gdpPercap)) + geom_jitter() + geom_smooth(method = )+ ggtitle("(
```



(c)

A 95% confidence interval for the growth rate of GDP per capita per year in the US is between \$125.3 and \$169.0.

```
confint(model1)
```

```
                 2.5 %       97.5 %
(Intercept) -327683.053 -241174.8803
year            125.323     169.0222
```