# ETC 2420/5242 Lab 5 2017

Di Cook

SOLUTION

Question 1 (1pt)

- Pick your favorite country (other than Australia). Find the parameter estimates from the `country_coefs` data frame. Do a hand-sketch of the fitted model. **Various answers, but should show the correct intercept and slope with axes marked and labels.**

Question 2 (5 pts, none for a)

a. Make a scatterplot of the linear model estimates for each country, slope vs intercept. Colour the points by continent.

```
ggplot(country_coefs, aes(x=intercept, y=year1950, colour=continent)) +
  geom_point()
```
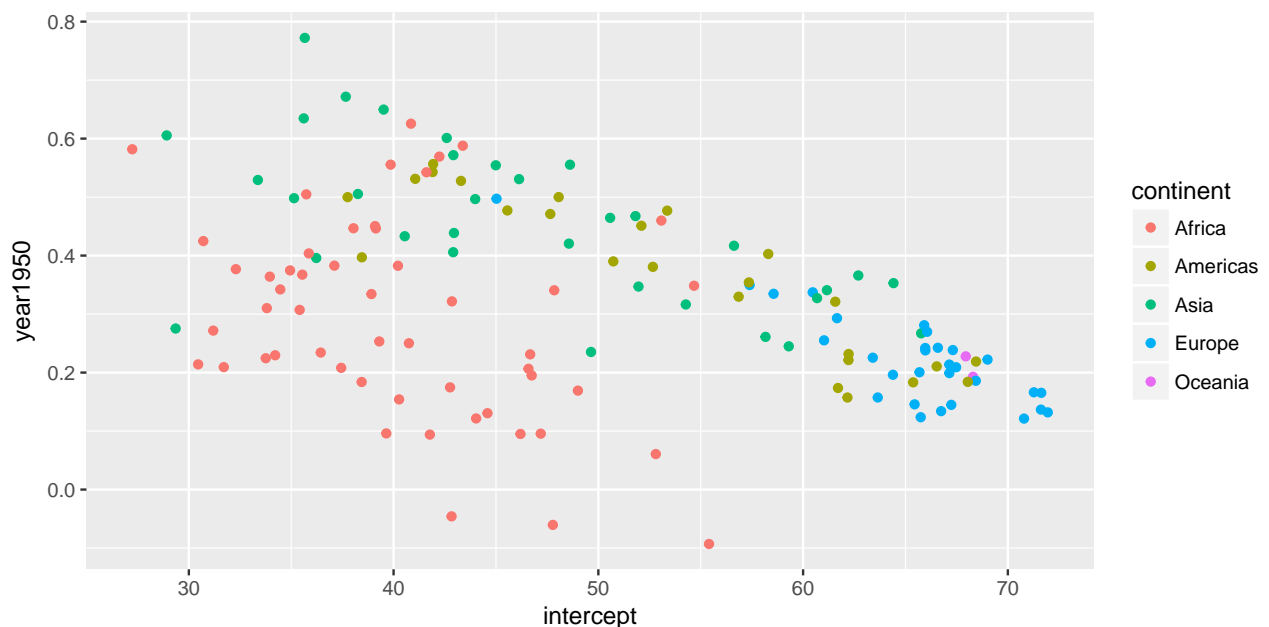

b. Statistically summarise the relationship between intercept and slope, using words like no association, positive linear association, negative linear association, weak, moderate, strong, outliers, clusters. **The association is negative moderate and linear, with some clustering by continent.**

c. Do you see a difference between continents? If so, explain what you see. **Africa shows the lowest intercepts and most variation in the slope. Europe is high on intercept and low on slope. Asia, like the Americas, is varied on intercept but relatively high on slope.**

d. What does it mean for a country to have a high intercept, e.g. 70? **The life expectancy in 1950 was quite high, e.g. 70 years.**

e. What does it mean for a country to have a high slope, e.g. 0.7? **The country had dramatic increase in life expectancy over the years. A value of 0.7 means life expectancy increased by 7 years for every decade.**

f. Make the plot interactive using the `plotly` package, and find out which countries had a negative slope.
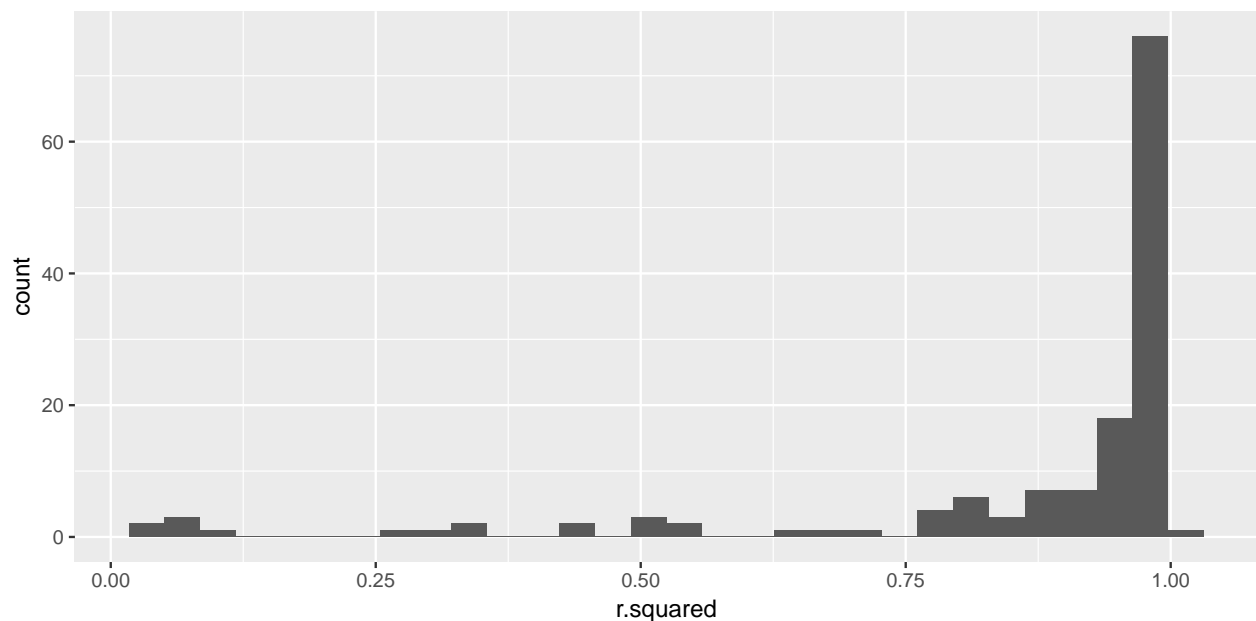Rwanda, Zambia, Zimbabwe

```
ggplot(country_coefs, aes(x=intercept, y=year1950,
                          colour=continent, label=country)) +
  geom_point()
```
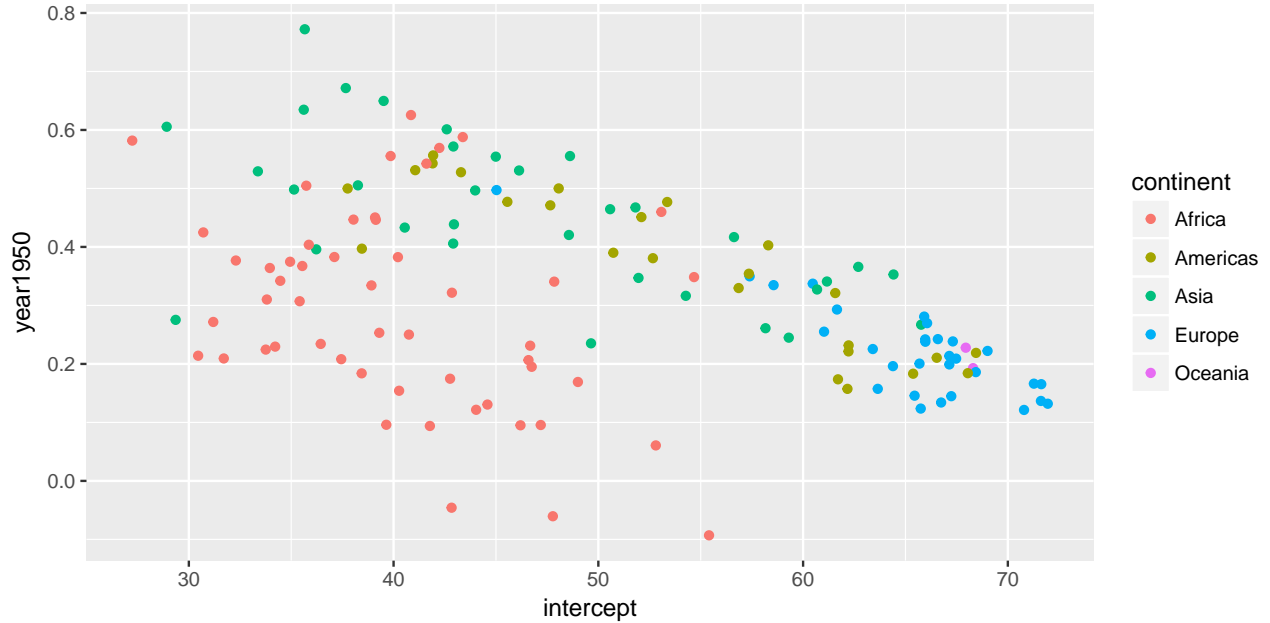


```
library(plotly)
ggplotly()
```

Question 3 (2pts, none for a)

a. Plot the R^2 values as a histogram.

b. Statistically describe the distribution using words like symmetric, skewed left, skewed right, unimodal, bimodal, multimodal, outliers. `The distributionn is skewed left, with a few really low values.`

c. What do you learn about the model fit across all 142 countries? `For most countries, the fit is very good. There are some countries where a linear model does not fit at all.`

Question 4 (2pts, none for a)

a. Examine the countries with the worst fit, countries with $R^2 < 0.45$, by making scatterplots of the data, with the linear model overlaid.
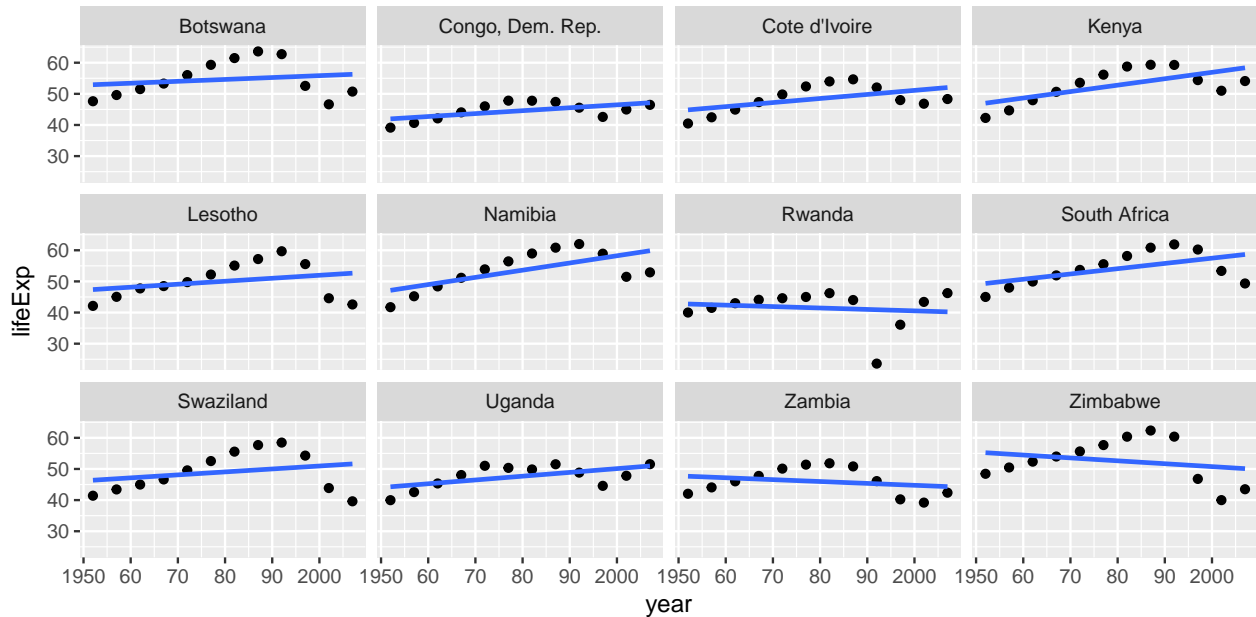
b. Each of these countries has a big dip in their life expectancy during the time of the study. Explain these using world history and current affairs information. (Feel free to google for news stories.)`Civil wars and AIDS`

c. Use the statistics to investigate something of your choice related to life expectancy across the world over those 50 years. `Various answers, could notice that the top R2 countries have very linear growth in life expectancy, or genocide in Cambodia, or country that had the most dramatic increase,...`