Statistics Project – Math 141

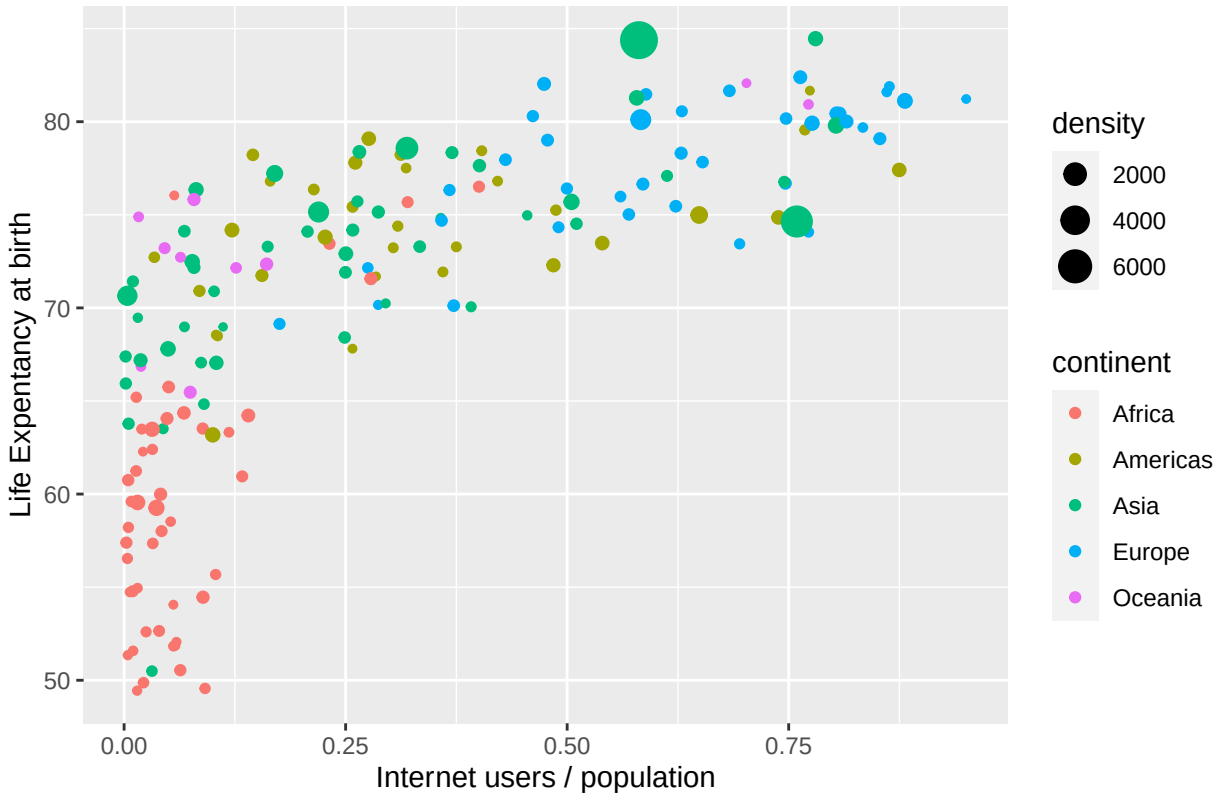
Dashiel Ward, Gavin Rimmer, Bailee Brunsmann, Harrison Nicholls

CIA Factbook

Question:

In a given nation, does the proportion of the population using the internet correlate to life expectancy at birth?

Figure 1



Dataset: What is it?

- Details on countries

Where does it come from?

- CIA Factbook

What is the description of each variable?

- See below

Variables:

Country - Categorical - Nominal Countries recognized by the CIA.

Area - Numerical - Continuous Land in Square km

Infant Mortality Rate - Numerical - Continuous Infant mortality rate compares the number of deaths of infants under one year old in a given year per 1,000 live births in the same year. This rate is often used as an indicator of the level of health in a country.

Population - Numerical - Discrete Population compares estimates from the US Bureau of the Census based on statistics from population censuses, vital statistics registration systems, or sample surveys pertaining to the recent past and on assumptions about future trends.

Population growth rate - Numerical - Continuous Population growth rate compares the average annual percent change in populations, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country. The rate may be positive or negative.

Birth Rate - Numerical - Continuous Birth rate compares the average annual number of births during a year per 1,000 persons in the population at midyear; also known as crude birth rate.

Death rate - Numerical - Continuous Death rate compares the average annual number of deaths during a year per 1,000 population at midyear; also known as crude death rate.

Net migration rate - Numerical - Continuous Net Migration rate compares the difference between the number of persons entering and leaving a country during the year per 1,000 persons (based on midyear population).

Maternal mortality rate – Numerical - Continuous The Maternal mortality rate (MMR) is the annual number of female deaths per 100,000 live births from any cause related to or aggravated by pregnancy or its management (excluding accidental or incidental causes).

Life expectancy at birth – Numerical - Discrete Life expectancy at birth compares the average number of years to be lived by a group of people born in the same year, if mortality at each age remains constant in the future. Life expectancy at birth is also a measure of overall quality of life in a country and summarizes the mortality at all ages.

Internet users – Numerical - Discrete Internet users compares the number of users within a country that access the Internet. Statistics vary from country to country and may include users who access the Internet at least several times a week to those who access it only once within a period of several months.

Further Research:

Lee, Cheng-Wen, The Relationship between Internet Environment and Life Expectancy in Asia

Article explores this question specifically in Asia, with relevant take-away in emphasising the disparity between countries with advanced telecommunication services and those without in regards to their general economic development in a globalized market.

Alzaid, Ahmed, Musleh Alsulami, Komal Komal, Adel-Maraghi, Examining the Relationship between the Internet and Life Expectancy

Article explores this question globally, finding that the economic development of a country is greatly bolstered by internet development, and has both direct and indirect impacts on the average life expectancy of its citizens.

Exploratory plots:

```
## Rows: 177
## Columns: 14
## $ country                   <chr> "Russia", "Canada", "United States", "China"~
## $ area                      <dbl> 17098242, 9984670, 9826675, 9596960, 8514877~
## $ birth_rate                <dbl> 11.87, 10.29, 13.42, 12.17, 14.72, 12.19, 19~
## $ death_rate                <dbl> 13.83, 8.31, 8.15, 7.44, 6.54, 7.07, 7.35, 7~
## $ infant_mortality_rate     <dbl> 7.08, 4.71, 6.17, 14.79, 19.21, 4.43, 43.19,~
## $ internet_users            <dbl> 40853000, 26960000, 245000000, 389000000, 75~
## $ life_exp_at_birth         <dbl> 70.16, 81.67, 79.56, 75.15, 73.28, 82.07, 67~
## $ maternal_mortality_rate   <dbl> 34, 12, 21, 37, 56, 7, 200, 77, 51, 540, 24,~
## $ net_migration_rate        <dbl> 1.69, 5.66, 2.45, -0.32, -0.15, 5.74, -0.05,~
## $ population                <dbl> 142470272, 34834841, 318892103, 1355692576, ~
## $ population_growth_rate    <dbl> -0.03, 0.76, 0.77, 0.44, 0.80, 1.09, 1.25, 0~
## $ continent                 <chr> "Europe", "Americas", "Americas", "Asia", "A~
## $ density                   <dbl> 8.332451, 3.488832, 32.451679, 141.262710, 2~
## $ internet_usage_proportion <dbl> 0.286747540, 0.773937794, 0.768284939, 0.286~
```

Figure 2

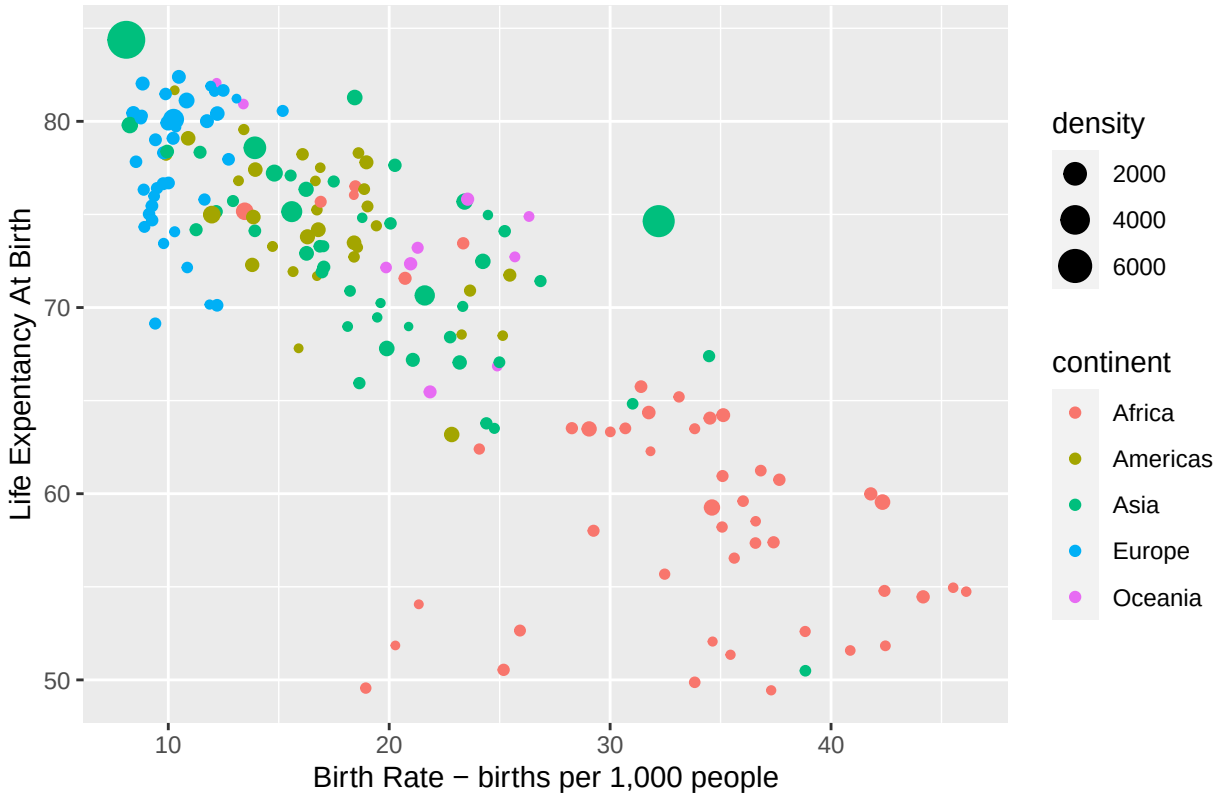


Figure 3

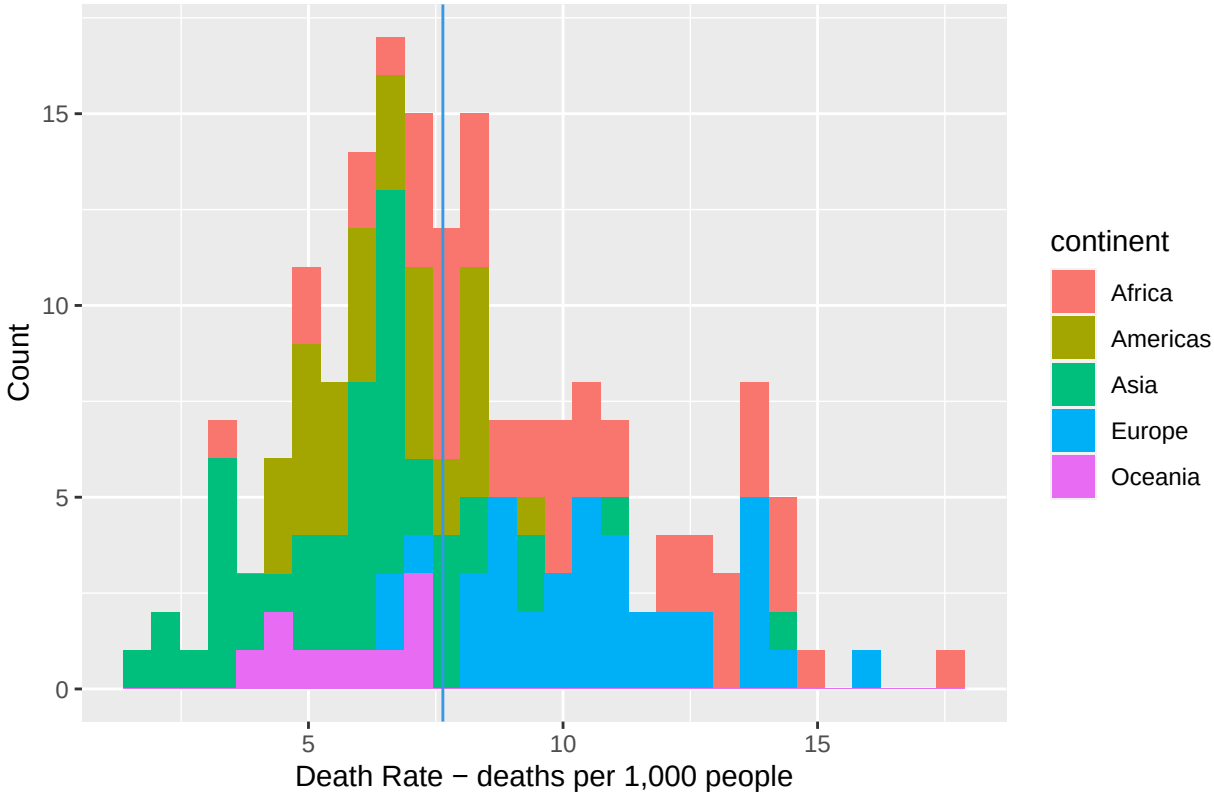


Figure 4

Statistical methods

```
continentmembercounts <- cia %>%
  group_by(continent) %>%
  count()

groups <- cia %>%
  group_by(continent) %>%
  select(birth_rate) %>%
  summarize(
    birth_rate = mean(birth_rate)) %>%
  mutate(
    difference_from_mean = birth_rate - mean(cia$birth_rate)
      )
```

Figure 5

```
## Adding missing grouping variables: `continent`
```

```
groups
```

```
## # A tibble: 5 x 3
##   continent birth_rate difference_from_mean
##   <chr>          <dbl>                <dbl>
## 1 Africa          32.2                11.6
## 2 Americas        16.9                -3.66
## 3 Asia            19.4                -1.14
## 4 Europe          10.4               -10.2
## 5 Oceania         21.0                 0.419
```

```
lm_mod <- lm(life_exp_at_birth ~ internet_usage_proportion, data = cia)
summary(lm_mod)
```

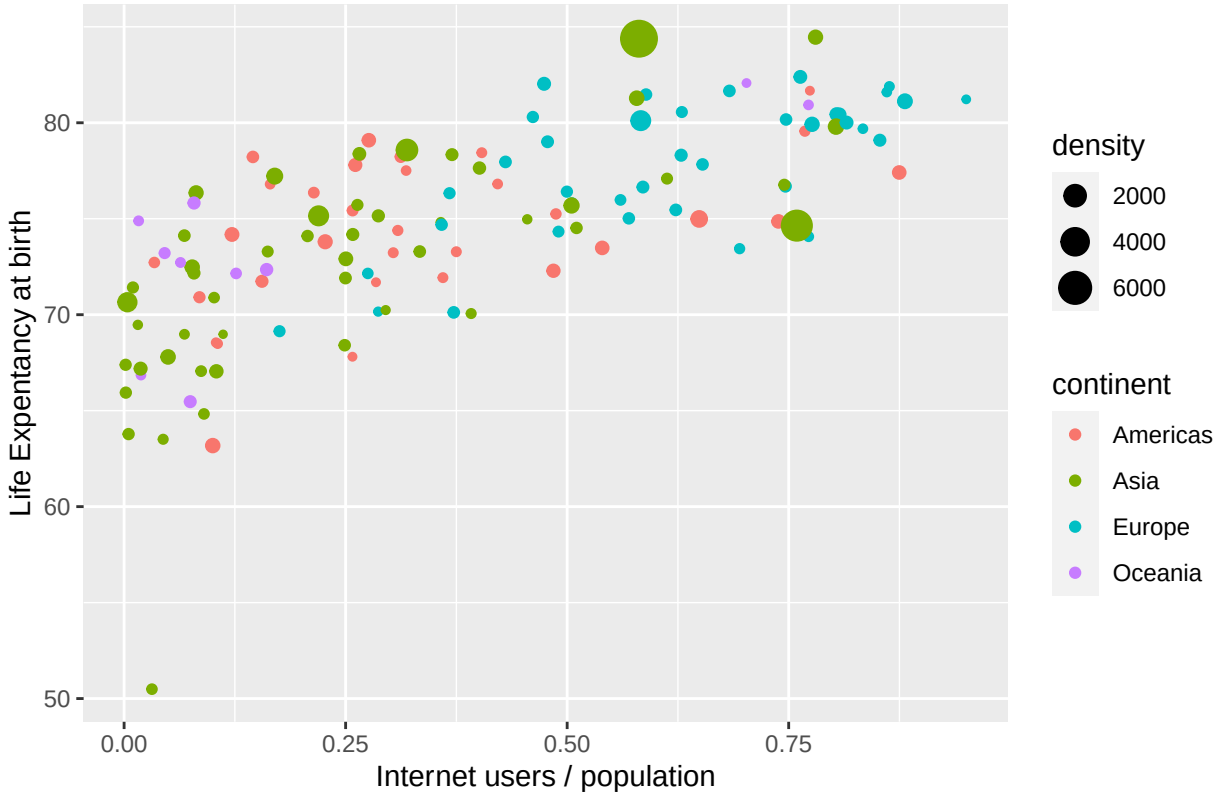
Linear model with Africa included

```
##
## Call:
## lm(formula = life_exp_at_birth ~ internet_usage_proportion, data = cia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.2903  -2.8736   0.1377   4.4186  11.1027
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 63.7069     0.6732   94.63   <2e-16 ***
## internet_usage_proportion  23.4594     1.6731   14.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.066 on 175 degrees of freedom
## Multiple R-squared:  0.5291, Adjusted R-squared:  0.5264
## F-statistic: 196.6 on 1 and 175 DF,  p-value: < 2.2e-16
```

```
noafrica <- cia %>%
  filter(continent != "Africa")

ggplot (data=noafrica, aes(x=internet_usage_proportion, y=life_exp_at_birth)) + geom_point(aes(color = c
```

Figure 6 − Africa Excluded




```
lm_mod_noafrica <- lm(life_exp_at_birth ~ internet_usage_proportion, data=noafrica)
summary(lm_mod_noafrica)
```

```
##
## Call:
## lm(formula = life_exp_at_birth ~ internet_usage_proportion, data = noafrica)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.4508  -2.3693   0.0082   2.6411   7.0030
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 69.5157     0.5602  124.08   <2e-16 ***
## internet_usage_proportion   13.5291     1.2007   11.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.678 on 127 degrees of freedom
## Multiple R-squared:  0.4999, Adjusted R-squared:  0.496
## F-statistic:   127 on 1 and 127 DF,  p-value: < 2.2e-16
```

Statistical Methods:

In performing an observational study, we are limited in the viability of performing randomization tests on our data. We can find correlations in our plots using R^2 and find the differences in means between different categorical groups, in our case Continents seems apt for a geopolitical analysis of the data. However, given that our sample population is itself the true population of all countries, there are no methods by which we can fabricate more samples via bootstrapping.

So as to glean a better understanding of the data observed, several steps were taken to parse from it usable information. These included:

Calculating the R^2: 0.5

Finding the mean birth rate per continent: (See Figure Six)

Finding the difference from the mean of the mean birth rates per continent: (See Figure Six)

Analysis of Data Visualizations:

Figure One: The proportion of internet users in the population of a given country correlates positively with life expectancy at birth. As indicated by the shading — indicating continent — these trends are not evenly dispersed across the nations of the world. Africa is clustered largely around the bottom left, with lower internet usership and life expectancy, while Europe is largely clustered in the top right, with higher internet usership and life expectancy. Oceania, Asia and the Americas are distributed across the plot.

Figure Two: Birth rate appears to be inversely proportional to life expectancy at birth, meaning that individuals in countries with higher life expectancy tend to have fewer children. As indicated by the shading — indicating continent — these trends are not evenly dispersed across the nations of the world. Africa and Europe again find themselves at opposing ends of the plot, with higher birth rates, lower life expectancy, and lower birth rates, higher life expectancy respectively. Once more, Oceania, Asia, and the Americas lie dispersed in the middle.

Figure Three: Africa and Europe take the upper end of our death rate histogram, with Asia, Oceania, and the Americas occupying the lower end. Africa seems to have the largest range in death rate. The graph takes a minor lower skew, though it makes an asymmetrical peak in greatest frequency around 8 deaths per 1,000 population.

Figure Four: Africa takes most of the lower end of our life expectancy histogram, while the Americas, Asia, Europe, and Oceania carry similar ranges, with Asia as the lower of the bunch and Europe as the upper of the four. Life expectancy seems to have an upper skew across all countries, with the highest frequency occurring around the 75-76 range.

Figure Six: In excluding Africa, we are provided with a new angle on Figure One. While the relationship is the same, the correlation appears to be less strong. The bias seen in Figure One also appears lower.