Rosling Demo

kvlinden

This R-Markdown file reproduces the code demoed in the class demo slideshow.

First, we load the libraries with the required programming tools and the raw data.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gapminder)
```

Here's the data (from the gapminder library).

```
gapminder
```

```
## # A tibble: 1,704 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 1,694 more rows
```

This computes the average life expectancy in the data over all countries and all times.

```r
gapminder %>%
  summarize(AvgLifeExp = mean(lifeExp))
```

```
## # A tibble: 1 x 1
##   AvgLifeExp
##        <dbl>
## 1       59.5
```

This groups the data by year and computes an average per group.

```r
gapminder %>%
  group_by(year) %>%
  summarize(AvgLifeExp = mean(lifeExp))
```

```
## # A tibble: 12 x 2
##     year AvgLifeExp
##    <int>      <dbl>
##  1  1952       49.1
##  2  1957       51.5
##  3  1962       53.6
##  4  1967       55.7
##  5  1972       57.6
##  6  1977       59.6
##  7  1982       61.5
##  8  1987       63.2
##  9  1992       64.2
## 10  1997       65.0
## 11  2002       65.7
## 12  2007       67.0
```

This takes the same data, grouped by year, and plots the average over time using a line plot.

```r
gapminder %>%
  group_by(year) %>%
  summarize(AvgLifeExp = mean(lifeExp)) %>%
  ggplot() + #<<
    aes(x=year, y=AvgLifeExp) +  #<<
    geom_line() #<<
```
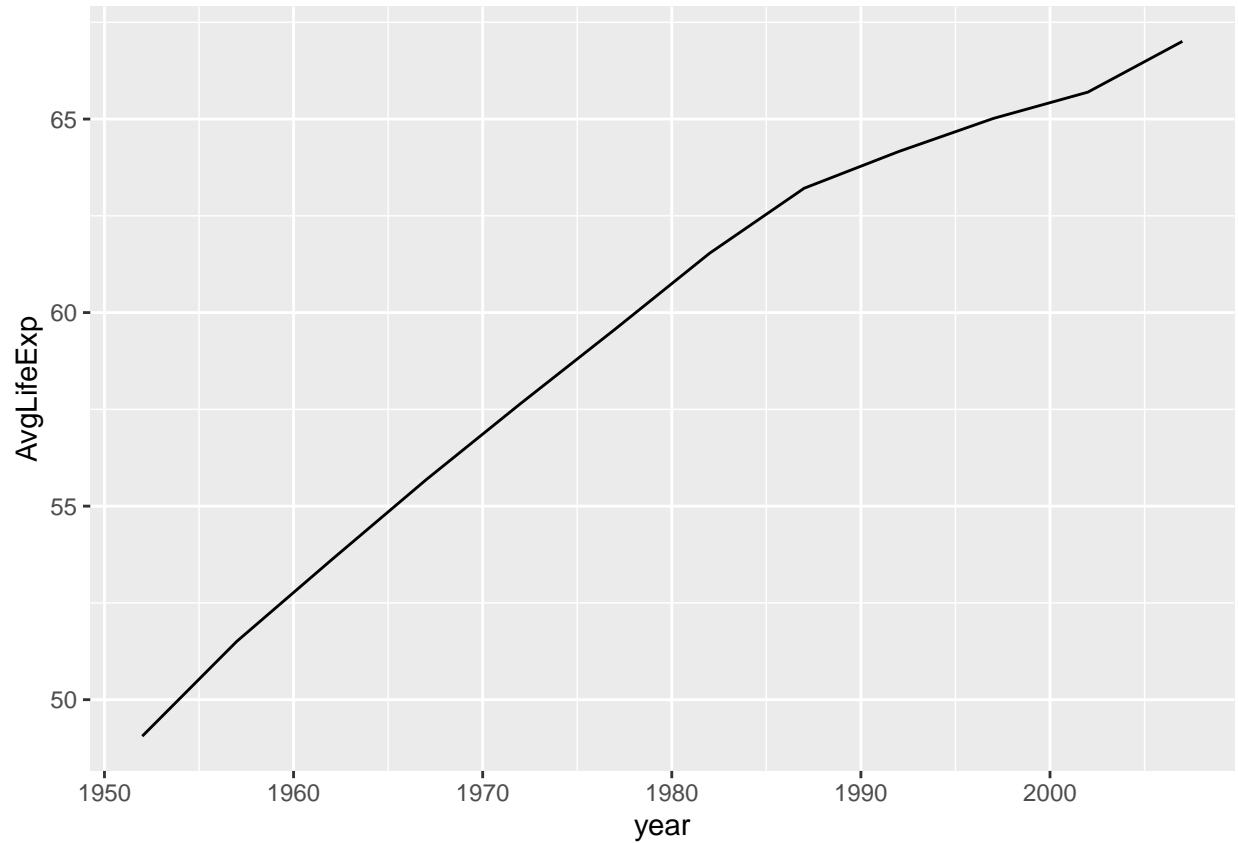
Finally, this graph approximates one time-stop of Rosling's famous animated plot (see: The Best Stats You've Ever Seen.

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
gapminder %>%
  filter(year == 2007) %>%
  ggplot() +
    aes(x = gdpPercap, y = lifeExp) +
    geom_point(alpha = .8) +
    aes(color = continent) +
    aes(size = pop) +
    scale_x_continuous(
      breaks = c(400, 4000, 40000),
```

```
    trans = "log10") +
labs(x = "GDP per Capita") +
labs(y = "Life Expectancy (years)") +
labs(color = "Continent") +
labs(size = "Population") +
scale_size_area(labels=label_comma()) +
theme_bw() +
annotation_logticks(sides = "b")
```
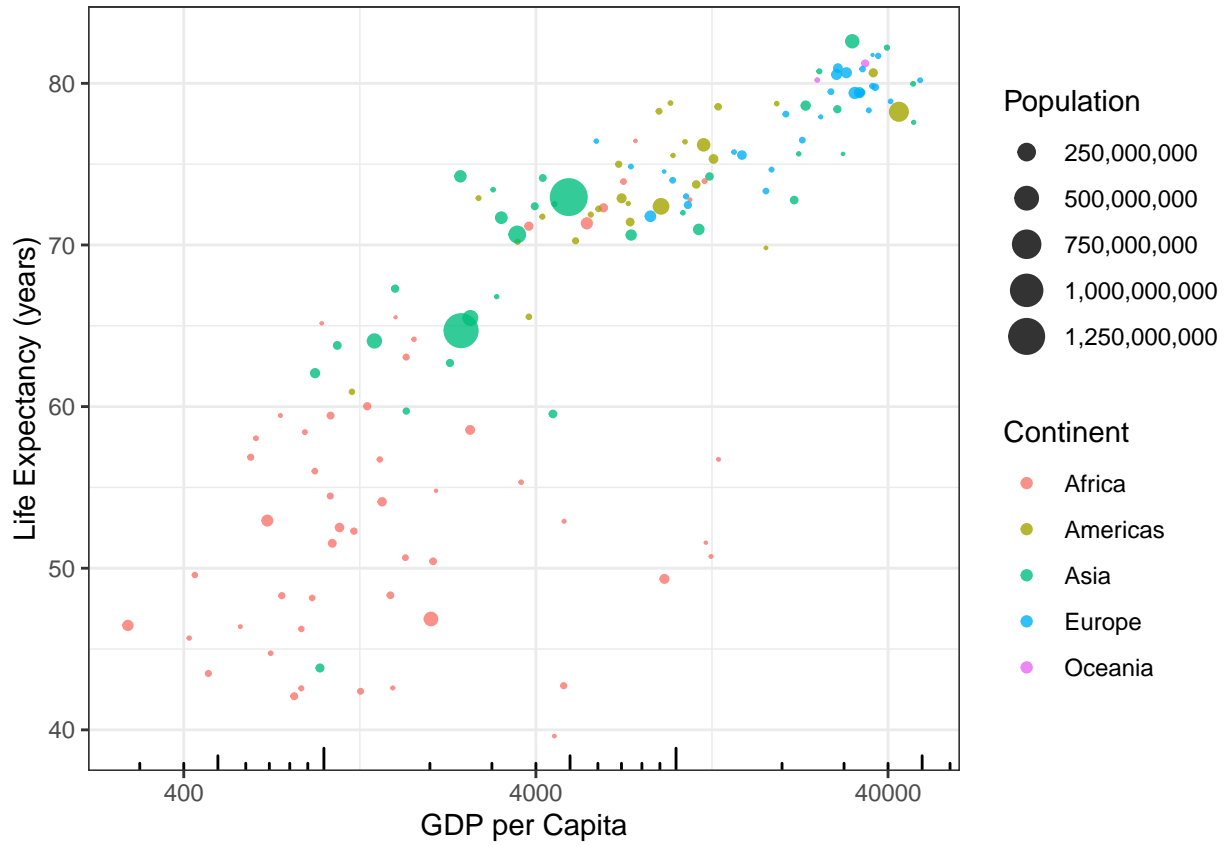

Question: Could you do this in Excel?