

gapminder data

Evelyn Takyi

March 1, 2017

Load gapminder data

```
gapminder <- read.csv('gapminder-FiveYearData.csv')
head(gapminder)

##      country year      pop continent lifeExp gdpPercap
## 1 Afghanistan 1952  8425333      Asia  28.801  779.4453
## 2 Afghanistan 1957  9240934      Asia  30.332  820.8530
## 3 Afghanistan 1962 10267083      Asia  31.997  853.1007
## 4 Afghanistan 1967 11537966      Asia  34.020  836.1971
## 5 Afghanistan 1972 13079460      Asia  36.088  739.9811
## 6 Afghanistan 1977 14880372      Asia  38.438  786.1134

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Structure and summary of the gapminder data

```
str(gapminder)

## 'data.frame':   1704 obs. of  6 variables:
## $ country : Factor w/ 142 levels "Afghanistan",...: 1 1 1 1 1 1 1 1 1 ...
## $ year    : int  1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
## $ pop     : num  8425333 9240934 10267083 11537966 13079460 ...
## $ continent: Factor w/ 5 levels "Africa","Americas",...: 3 3 3 3 3 3 3 3 3 ...
## $ lifeExp  : num  28.8 30.3 32 34 36.1 ...
## $ gdpPercap: num  779 821 853 836 740 ...

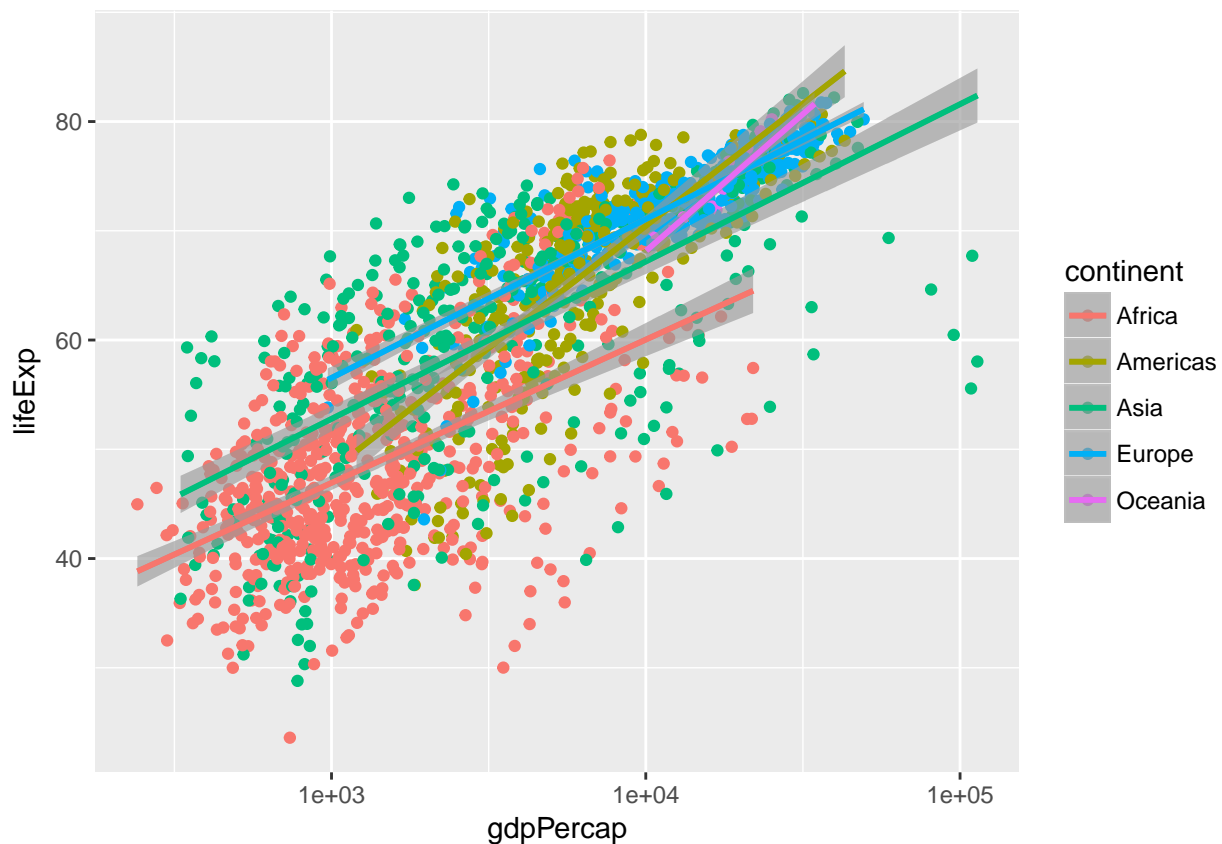
summary(gapminder)

##      country      year      pop      continent
## Afghanistan: 12   Min.   :1952   Min.   :6.001e+04   Africa :624
## Albania     : 12   1st Qu.:1966   1st Qu.:2.794e+06   Americas:300
```

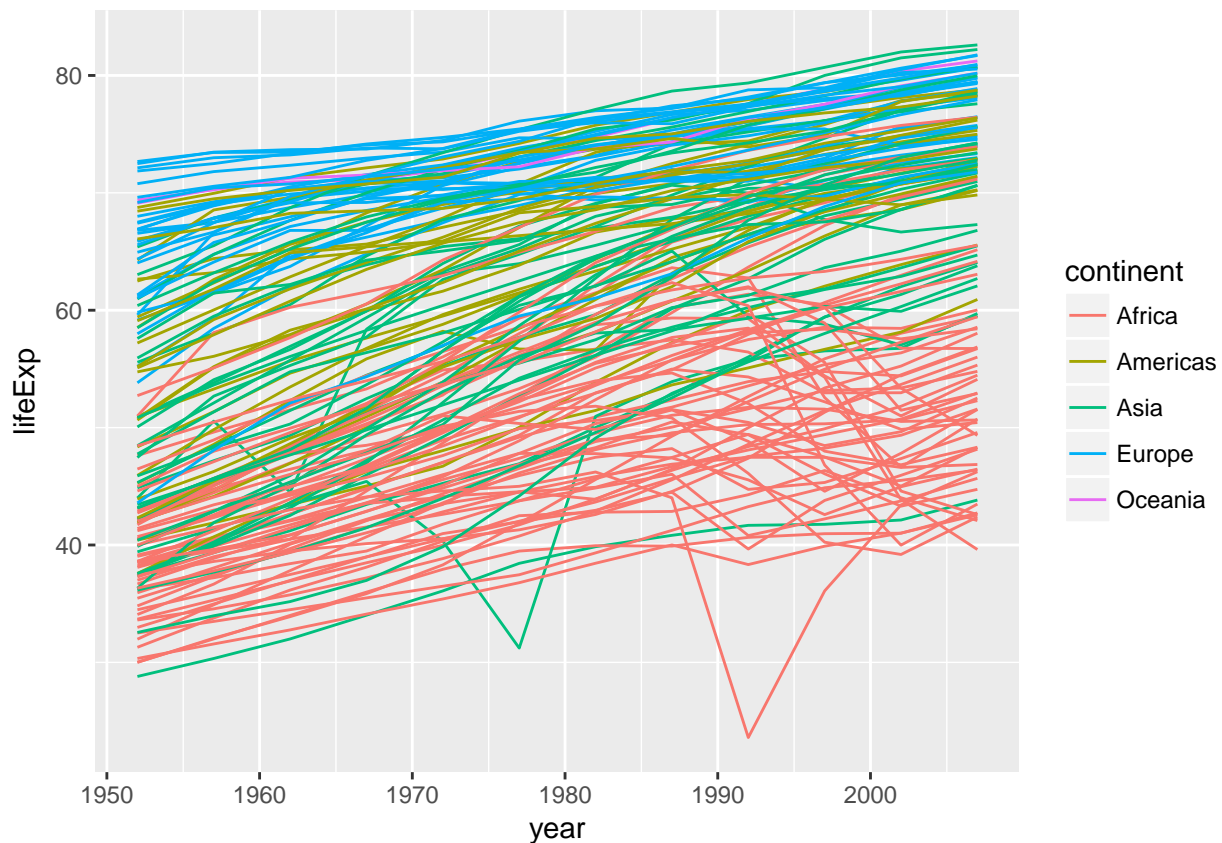
```
## Algeria      : 12   Median :1980   Median :7.024e+06   Asia      :396
## Angola       : 12   Mean    :1980   Mean    :2.960e+07   Europe    :360
## Argentina    : 12   3rd Qu.:1993   3rd Qu.:1.959e+07   Oceania   : 24
## Australia    : 12   Max.    :2007   Max.    :1.319e+09
## (Other)      :1632
##   lifeExp      gdpPercap
## Min.   :23.60   Min.    : 241.2
## 1st Qu.:48.20   1st Qu.: 1202.1
## Median :60.71   Median : 3531.8
## Mean   :59.47   Mean    : 7215.3
## 3rd Qu.:70.85   3rd Qu.: 9325.5
## Max.   :82.60   Max.    :113523.1
##
```

Plots

```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp, colour = continent)) +
  geom_point() + geom_smooth(method = 'lm') +
  scale_x_log10() + geom_smooth(method = 'lm')
```



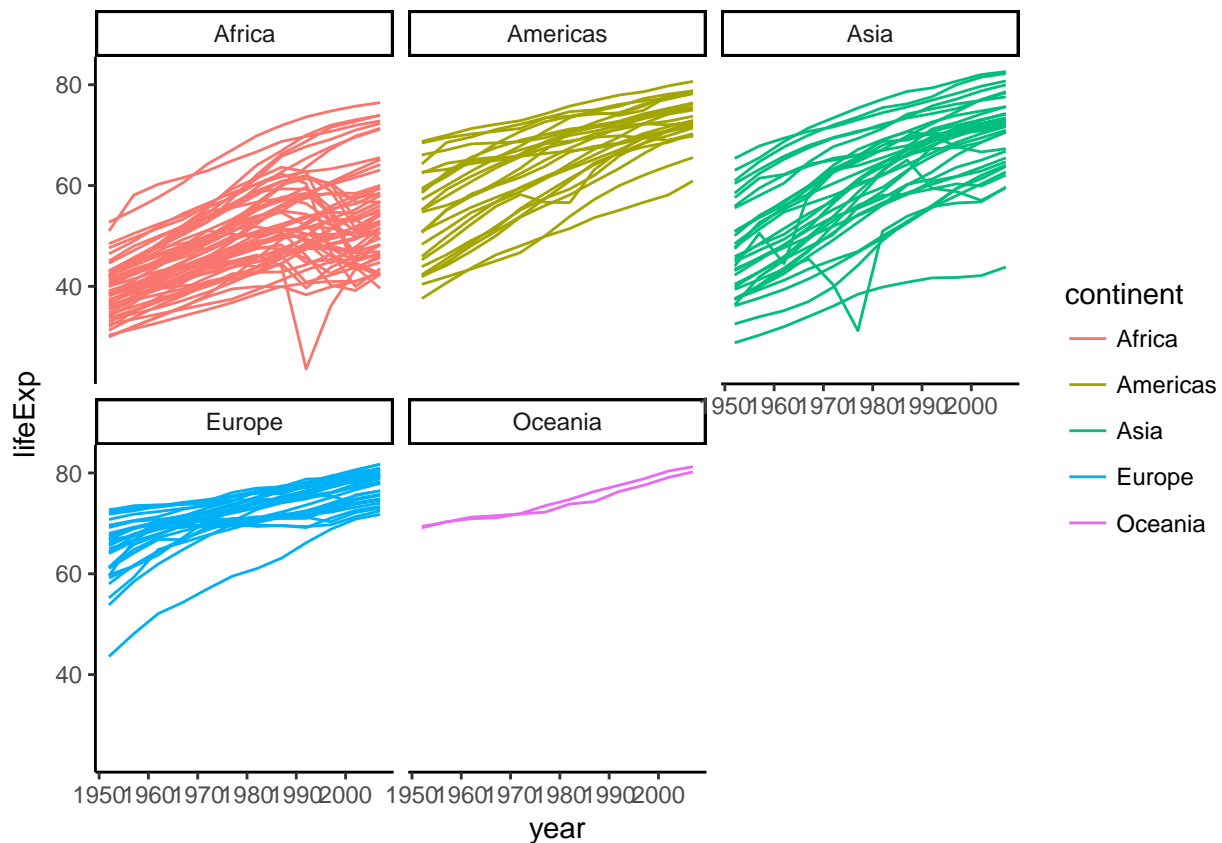
```
ggplot(gapminder, aes(x = year, y = lifeExp, group = country, width = 3.5, height = 2.5)) +
  geom_line(aes(colour = continent))
```



```
reg <- lm(lifeExp ~ gdpPercap, data=gapminder)
summary(reg)
```

```
##
## Call:
## lm(formula = lifeExp ~ gdpPercap, data = gapminder)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -82.754  -7.758   2.176   8.225  18.426
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.396e+01  3.150e-01  171.29  <2e-16 ***
## gdpPercap    7.649e-04  2.579e-05   29.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.49 on 1702 degrees of freedom
## Multiple R-squared:  0.3407, Adjusted R-squared:  0.3403
## F-statistic: 879.6 on 1 and 1702 DF, p-value: < 2.2e-16
```

```
ggplot(gapminder, aes(x = year, y = lifeExp, group = country, width = 3.5, height = 2.5)) +
  geom_line(aes(colour = continent)) + facet_wrap(~ continent) + theme_classic()
```



```
gre <- lm(lifeExp ~ year, data=gapminder)
summary(gre)
```

```
##
## Call:
## lm(formula = lifeExp ~ year, data = gapminder)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.949  -9.651   1.697  10.335  22.158
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -585.65219   32.31396  -18.12  <2e-16 ***
## year          0.32590    0.01632   19.96  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.63 on 1702 degrees of freedom
## Multiple R-squared:  0.1898, Adjusted R-squared:  0.1893
## F-statistic: 398.6 on 1 and 1702 DF,  p-value: < 2.2e-16
```