Statistics with R

Victoria

2024-12-11

#Statistical Analysis with R

Installing Dataset

```
install.packages("gapminder")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(gapminder)
data("gapminder")
```

Summarizing Data

```
summary(gapminder)
```

```
##         country        continent        year         lifeExp
##  Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60
##  Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20
##  Algeria    :  12   Asia    :396   Median :1980   Median :60.71
##  Angola     :  12   Europe  :360   Mean   :1980   Mean   :59.47
##  Argentina  :  12   Oceania : 24   3rd Qu.:1993   3rd Qu.:70.85
##  Australia  :  12                  Max.   :2007   Max.   :82.60
##  (Other)    :1632
##       pop              gdpPercap
##  Min.   :6.001e+04   Min.   :   241.2
##  1st Qu.:2.794e+06   1st Qu.:  1202.1
##  Median :7.024e+06   Median :  3531.8
##  Mean   :2.960e+07   Mean   :  7215.3
##  3rd Qu.:1.959e+07   3rd Qu.:  9325.5
##  Max.   :1.319e+09   Max.   :113523.1
##
```

```
mean(gapminder$gdpPercap)
```

```
## [1] 7215.327
```

```
median(gapminder$pop)
```

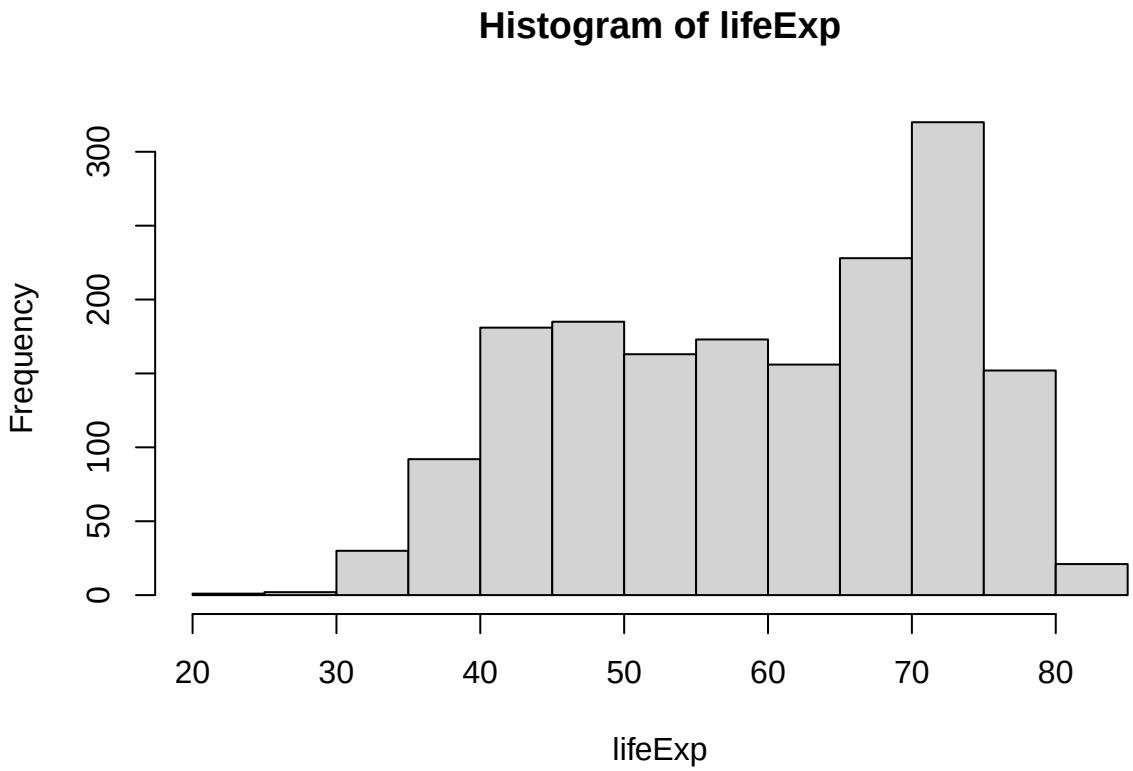
```
## [1] 7023596
```

Assigning Variables

```
x <- mean(gapminder$gdpPercap)
```

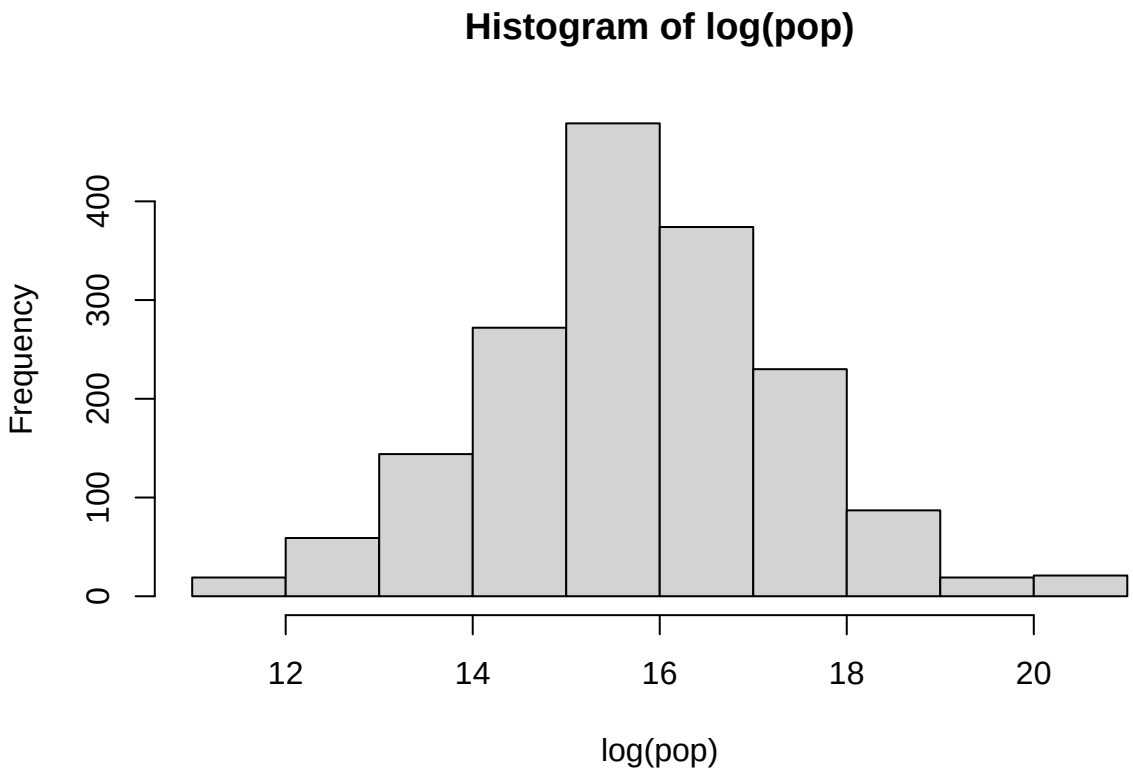
Data Distribution

```
attach(gapminder)
hist(lifeExp)
```



Histograms

```
hist(log(pop))
```



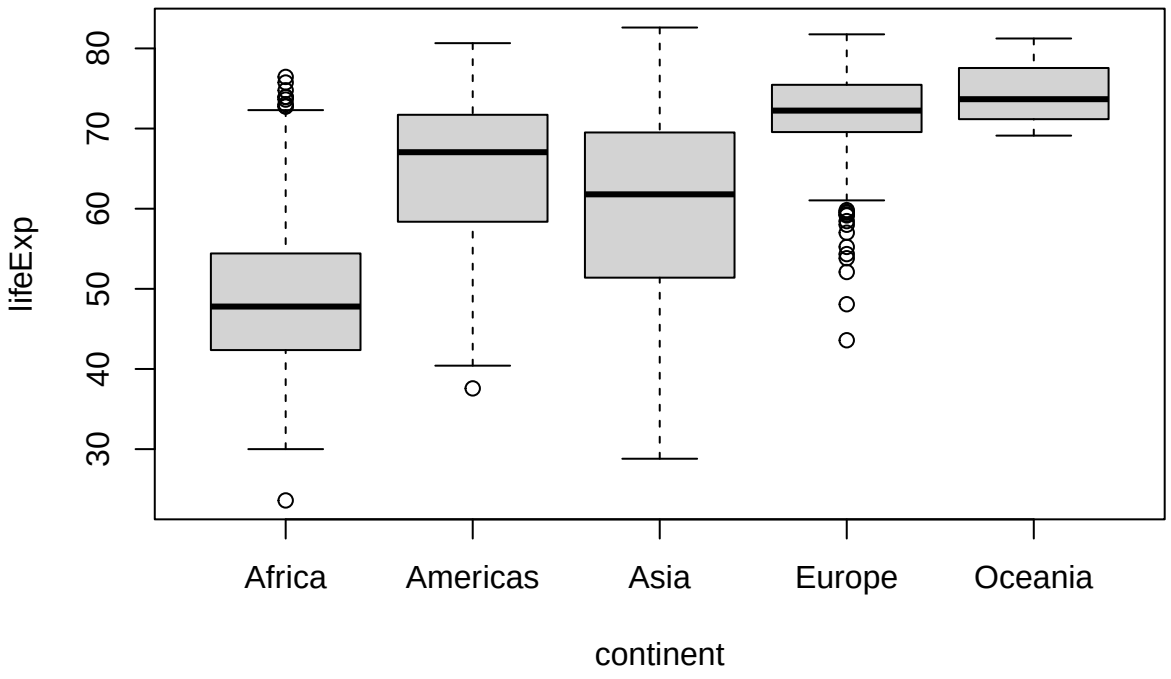
####

BoxPlots

```
attach(gapminder)
```

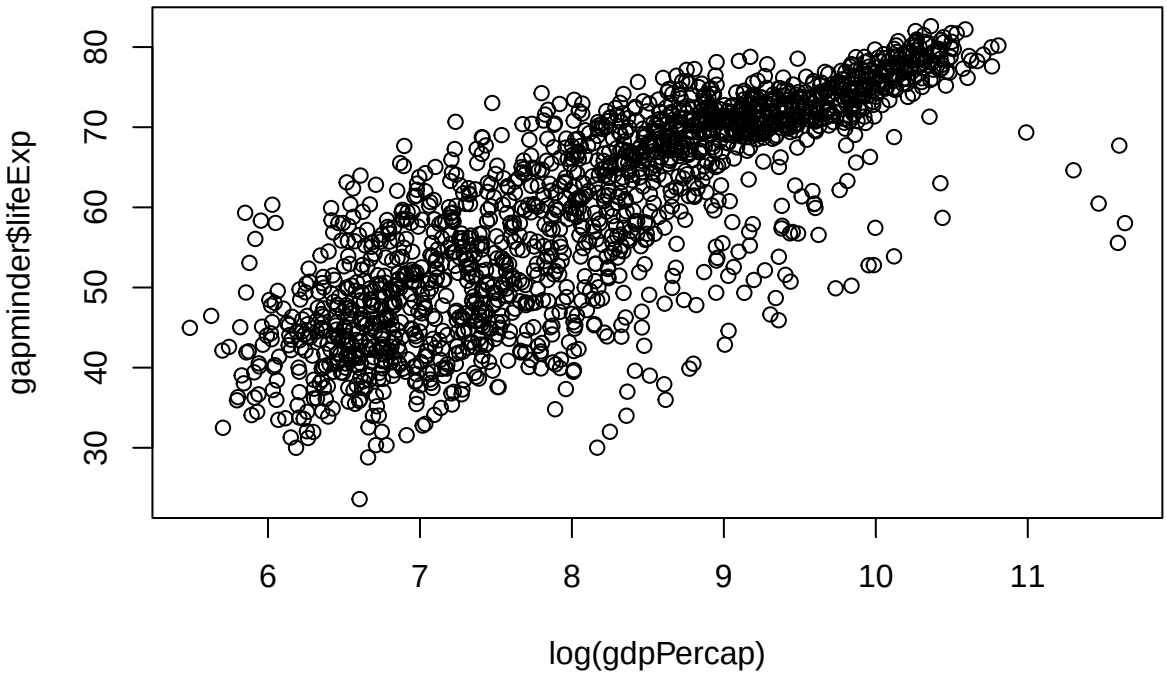
```
## The following objects are masked from gapminder (pos = 3):
##
##     continent, country, gdpPercap, lifeExp, pop, year
```

```
boxplot(lifeExp~continent)
```



Scatterplot

```
plot(gapminder$lifeExp~log(gdpPercap))
```



###

Data Manipulation with Dplyr

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
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
```

```
gapminder %>%
  select(country, lifeExp) %>%
  filter(country == "South Africa" |
           country == "Ireland") %>%
  group_by(country) %>%
  summarise(Average_Life = mean(lifeExp))
```

```
## # A tibble: 2 x 2
##   country      Average_Life
##   <fct>               <dbl>
## 1 Ireland              73.0
## 2 South Africa         54.0
```

```
install.packages("dplyr")
```

Performing Ttest

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(dplyr)

df1 <- gapminder %>%
  select(country, lifeExp) %>%
  filter(country == "South Africa" |
           country == "Ireland")

t.test(data = df1, lifeExp ~ country)
```

```
##
##  Welch Two Sample t-test
##
## data:  lifeExp by country
## t = 10.067, df = 19.109, p-value = 4.466e-09
## alternative hypothesis: true difference in means between group Ireland and group South Africa is not
## 95 percent confidence interval:
##  15.07022 22.97794
## sample estimates:
##      mean in group Ireland mean in group South Africa
##                   73.01725                   53.99317
```

Vizualization with ggplot2

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

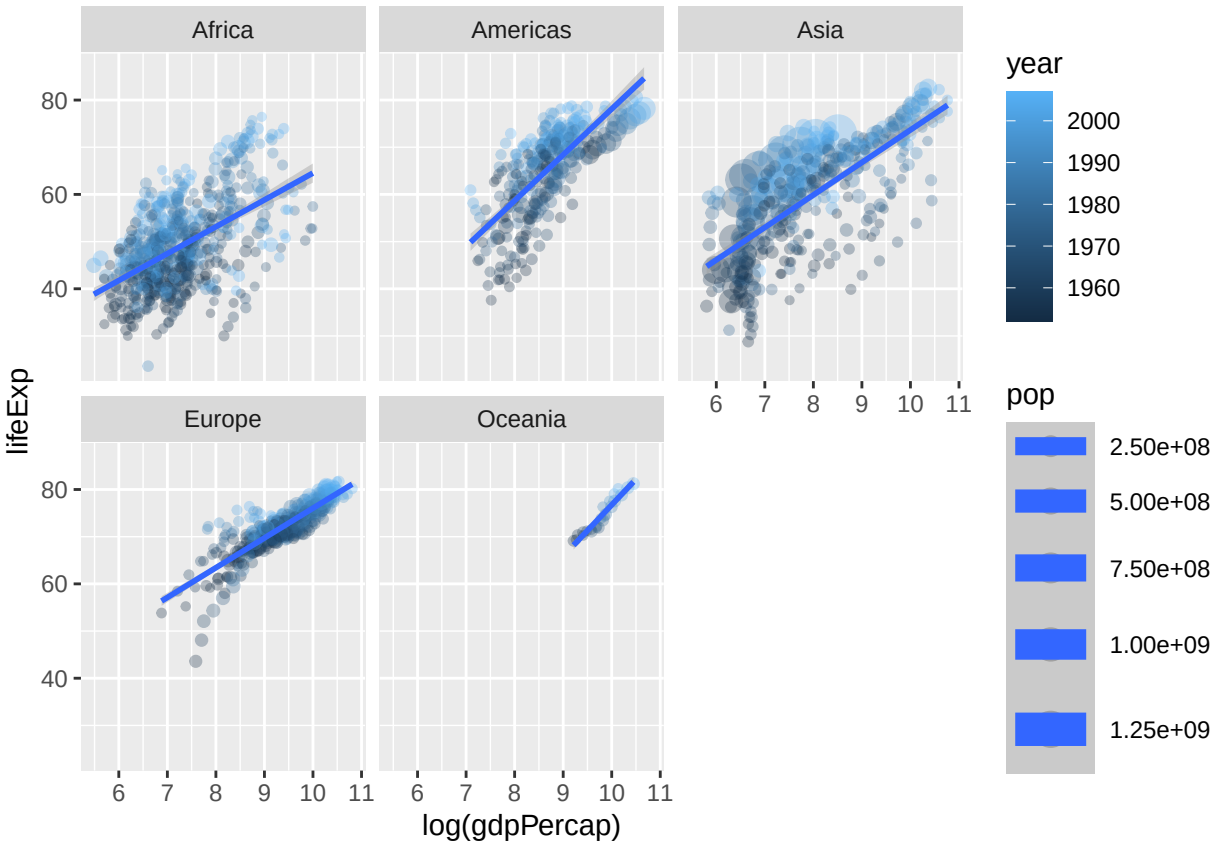
```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
library(dplyr)

gapminder %>%
  filter(gdpPercap < 50000) %>%
  ggplot(aes(x=log(gdpPercap), y=lifeExp, col=year, size= pop ))+
  geom_point(alpha=0.3)+
  geom_smooth(method = lm)+
  facet_wrap(~continent)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Linear Regression

```
attach(gapminder)
```

```
## The following objects are masked from gapminder (pos = 5):
##
##     continent, country, gdpPercap, lifeExp, pop, year
```

```
## The following objects are masked from gapminder (pos = 6):
##
##     continent, country, gdpPercap, lifeExp, pop, year
```

```
summary(lm(lifeExp ~ gdpPercap+pop))
```

```
##
## Call:
## lm(formula = lifeExp ~ gdpPercap + pop)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.754  -7.745   2.055   8.212  18.534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.365e+01  3.225e-01  166.36  < 2e-16 ***
## gdpPercap   7.676e-04  2.568e-05   29.89  < 2e-16 ***
## pop         9.728e-09  2.385e-09    4.08 4.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 10.44 on 1701 degrees of freedom
## Multiple R-squared:  0.3471, Adjusted R-squared:  0.3463
## F-statistic: 452.2 on 2 and 1701 DF,  p-value: < 2.2e-16
```