

task

2025-07-15

```
install.packages("gapminder")

## 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)

install.packages("ggplot2")

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

library(gapminder)
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

library(ggplot2)

# Subset data for 2007
data_2007 <- gapminder %>% filter(year == 2007, continent %in% c("Europe", "Asia"))

# Check sample size
table(data_2007$continent)

##
##   Africa Americas      Asia  Europe Oceania
##      0         0       33      30       0

# Check normality (Shapiro-wilk) - sample size is large, so approximation OK
by(data_2007$lifeExp, data_2007$continent, shapiro.test)

## data_2007$continent: Africa
## NULL
## -----
## data_2007$continent: Americas
## NULL
## -----
```

```

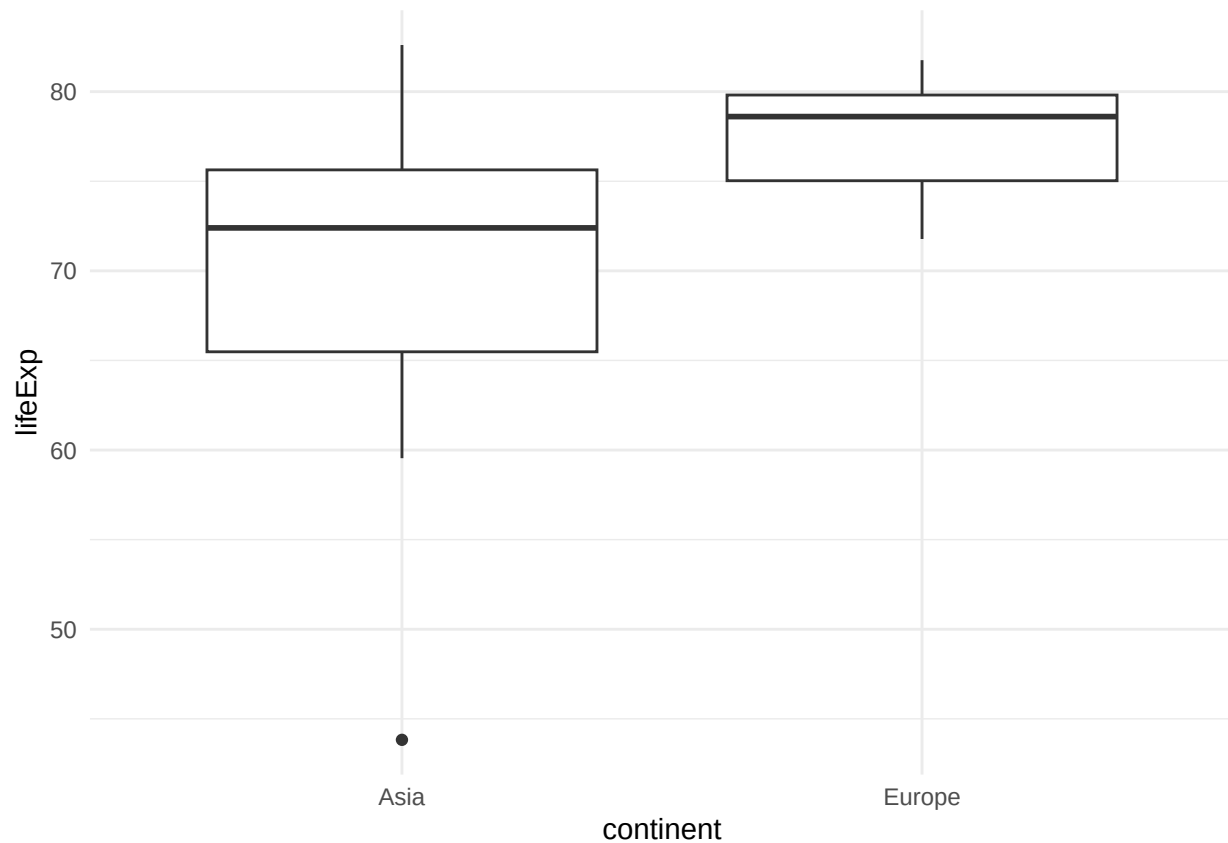
## data_2007$continent: Asia
##
## Shapiro-Wilk normality test
##
## data: dd[x, ]
## W = 0.92228, p-value = 0.02119
##
## -----
## data_2007$continent: Europe
##
## Shapiro-Wilk normality test
##
## data: dd[x, ]
## W = 0.92333, p-value = 0.03274
##
## -----
## data_2007$continent: Oceania
## NULL

# Check homogeneity of variances
var.test(lifeExp ~ continent, data = data_2007)

##
## F test to compare two variances
##
## data: lifeExp by continent
## F = 7.1426, num df = 32, denom df = 29, p-value = 7.374e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 3.439821 14.639586
## sample estimates:
## ratio of variances
## 7.14257

# Boxplot for visual inspection
ggplot(data_2007, aes(x = continent, y = lifeExp)) +
  geom_boxplot() +
  theme_minimal()

```



```
# Run t-test
t.test(lifeExp ~ continent, data = data_2007, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: lifeExp by continent
## t = -4.4802, df = 61, p-value = 3.34e-05
## alternative hypothesis: true difference in means between group Asia and group Europe is not equal to
## 95 percent confidence interval:
## -10.008747 -3.831484
## sample estimates:
## mean in group Asia mean in group Europe
## 70.72848 77.64860
```