

Untitled123

2025-06-25

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:

```
##
## The downloaded binary packages are in
## /var/folders/sm/l08k85g17pnf9wt_67pggsy00000gn/T//RtmpdmdZAE/downloaded_packages
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
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
```

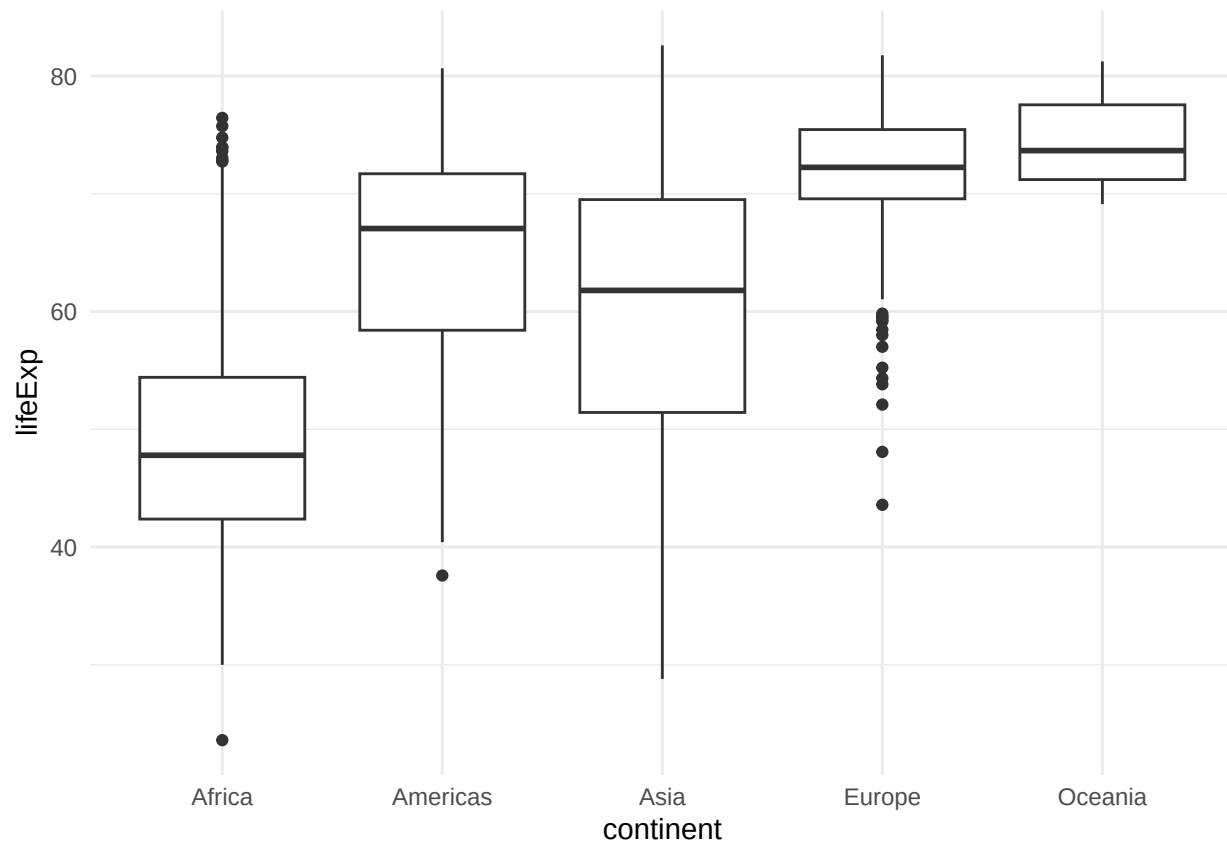
```
library(ggplot2)
data("gapminder")
head(gapminder)
```

```
## # A tibble: 6 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.
```

```
anova_model <- aov(lifeExp ~ continent, data = gapminder)
summary(anova_model)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## continent      4 139343   34836   408.7 <2e-16 ***
## Residuals    1699 144805      85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(gapminder, aes(x = continent, y = lifeExp)) +
  geom_boxplot() +
  theme_minimal()
```



```
TukeyHSD(anova_model)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lifeExp ~ continent, data = gapminder)
##
## $continent
##
```

	diff	lwr	upr	p adj
Americas-Africa	15.793407	14.022263	17.564550	0.0000000
Asia-Africa	11.199573	9.579887	12.819259	0.0000000
Europe-Africa	23.038356	21.369862	24.706850	0.0000000
Oceania-Africa	25.460878	20.216908	30.704848	0.0000000
Asia-Americas	-4.593833	-6.523432	-2.664235	0.0000000
Europe-Americas	7.244949	5.274203	9.215696	0.0000000
Oceania-Americas	9.667472	4.319650	15.015293	0.0000086
Europe-Asia	11.838783	10.002952	13.674614	0.0000000
Oceania-Asia	14.261305	8.961718	19.560892	0.0000000
Oceania-Europe	2.422522	-2.892185	7.737230	0.7250559

If both the lower bound (lwr) and upper bound (upr) do not include 0 (i.e., the confidence interval does not cross 0), and the adjusted p-value (p adj) is less than 0.05, it indicates a statistically significant difference in GDP per capita between the two continents.

If the confidence interval includes 0 (e.g., Asia-Americas and Oceania-Europe), it indicates that the difference is not statistically significant.

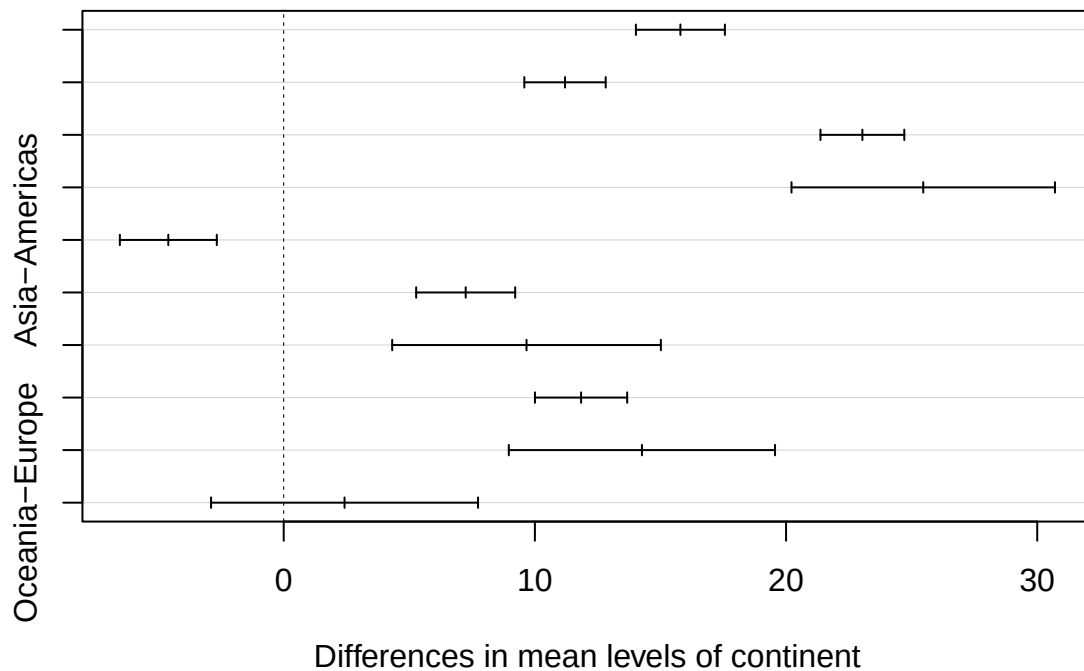
The GDP differences between Africa and all other continents are statistically significant.

The differences between Europe and both Asia and the Americas are also significant.

However, the differences between Asia and the Americas, as well as between Oceania and Europe, are not statistically significant.

```
plot(TukeyHSD(anova_model))
```

95% family-wise confidence level



```
library(gapminder)
library(dplyr)
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
```

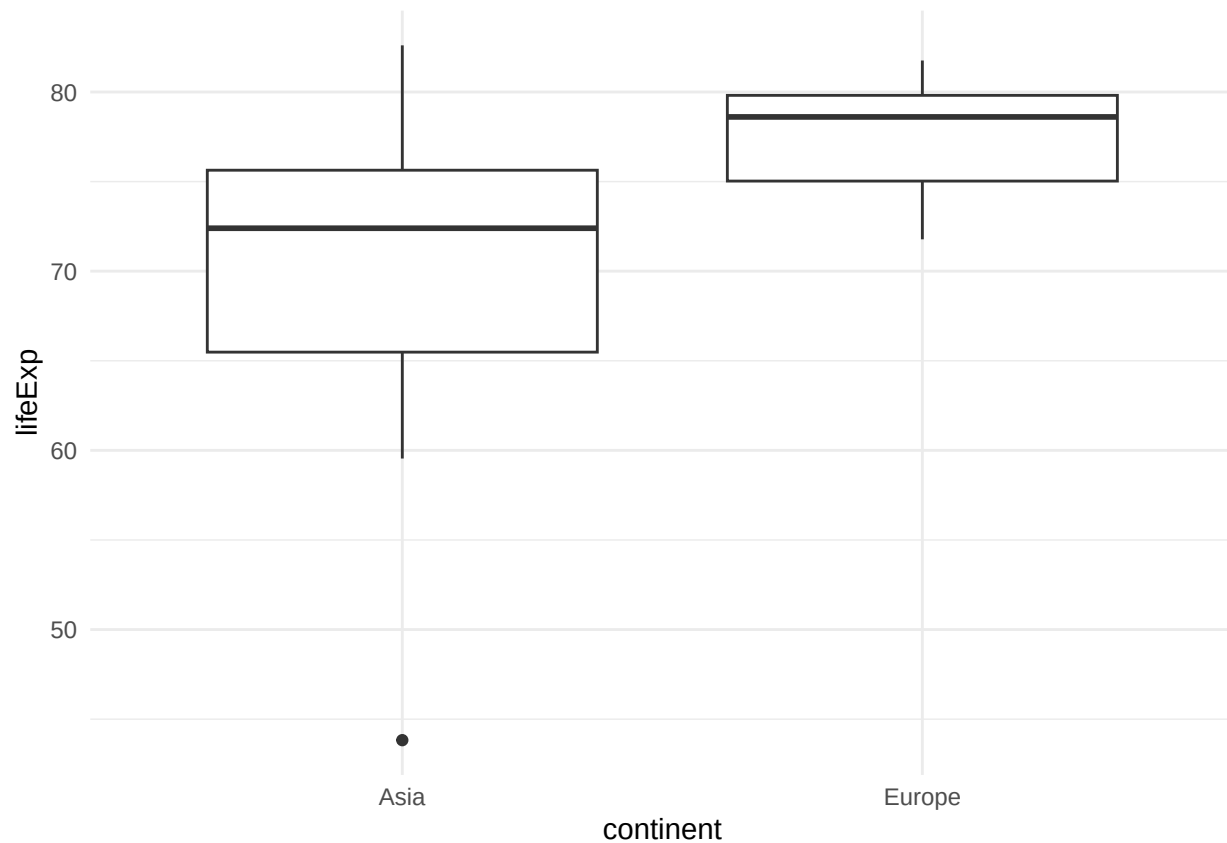
```
# Check normality (Shapiro-Wilk) - sample size is large, so approximation OK
by(data_2007$lifeExp, data_2007$continent, shapiro.test) #whether is normal distribution #eu and asia l
```

```
## 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
```

```
ggplot(data_2007, aes(x = continent, y = lifeExp)) +
  geom_boxplot() +
  theme_minimal()
```



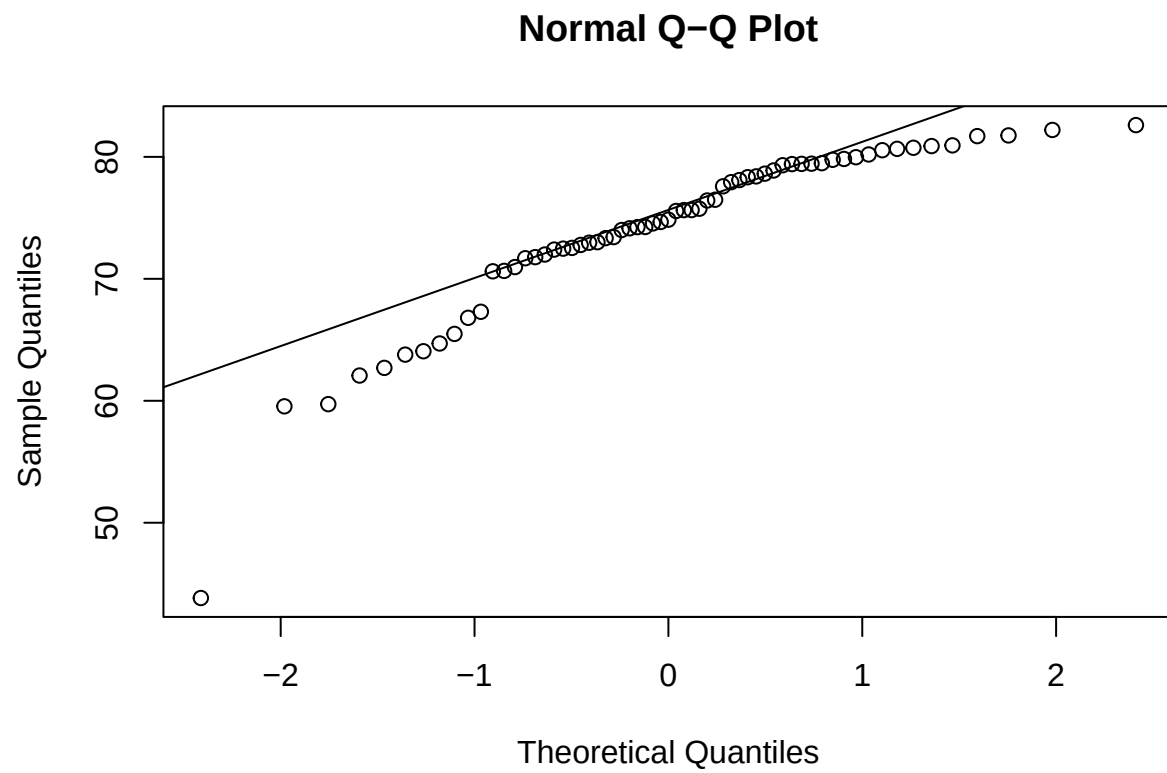
```
t.test(lifeExp ~ continent, data = data_2007, var.equal = TRUE)
```

```
##
## 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 0
## 95 percent confidence interval:
## -10.008747 -3.831484
## sample estimates:
## mean in group Asia mean in group Europe
## 70.72848 77.64860
```

```
wilcox.test(lifeExp ~ continent, data = data_2007)
```

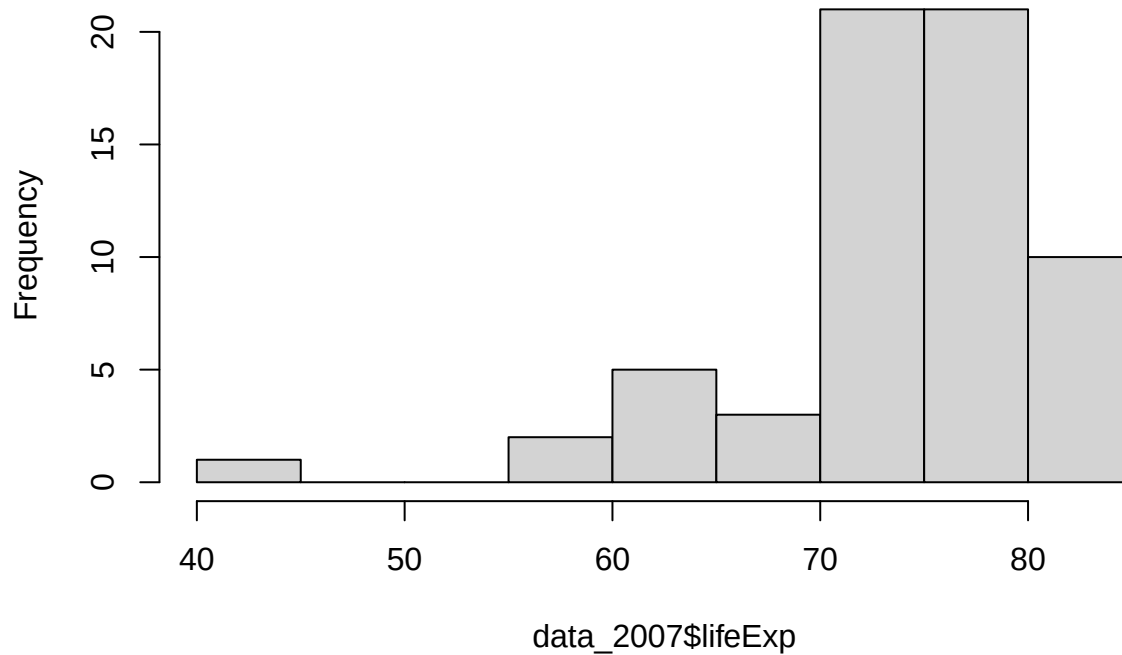
```
##
## Wilcoxon rank sum exact test
##
## data: lifeExp by continent
## W = 196, p-value = 1.944e-05
## alternative hypothesis: true location shift is not equal to 0
```

```
qqnorm(data_2007$lifeExp); qqline(data_2007$lifeExp)
```



```
hist(data_2007$lifeExp)
```

Histogram of data_2007\$lifeExp



```
shapiro.test(data_2007$lifeExp)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: data_2007$lifeExp  
## W = 0.86064, p-value = 4.069e-06
```

```
subset_data <- gapminder %>% filter(continent %in% c("Europe", "Asia"))  
t.test(lifeExp ~ continent, data = subset_data)
```

```
##  
## Welch Two Sample t-test  
##  
## data: lifeExp by continent  
## t = -17.899, df = 565.16, p-value < 2.2e-16  
## alternative hypothesis: true difference in means between group Asia and group Europe is not equal to  
## 95 percent confidence interval:  
## -13.13792 -10.53965  
## sample estimates:  
## mean in group Asia mean in group Europe  
## 60.06490 71.90369
```

```
anova_model <- aov(lifeExp ~ continent, data = gapminder)
summary(anova_model)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## continent      4 139343    34836   408.7 <2e-16 ***
## Residuals    1699 144805         85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
kruskal.test(lifeExp ~ continent, data = gapminder)
```

```
##
## Kruskal-Wallis rank sum test
##
## data:  lifeExp by continent
## Kruskal-Wallis chi-squared = 843.38, df = 4, p-value < 2.2e-16
```