

week7_Activity

2025-06-25

```
install.packages("gapminder")

## 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)

install.packages("dplyr")

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

install.packages("car")

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

install.packages("multcomp")

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

install.packages("ggpubr")

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

install.packages("broom")

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

library(gapminder)
library(ggplot2)
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(car)
```

```

## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##      recode
library(multcomp)

## Loading required package: mvtnorm
## Loading required package: survival
## Loading required package: TH.data
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##      select
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##      geyser
library(ggpubr)
library(broom)

data("gapminder")
# Ensure dplyr is loaded
library(dplyr)

# Use dplyr::select to avoid masking issues
gap_2007 <- gapminder %>%
  filter(year == 2007) %>%
  dplyr::select(continent, gdpPercap)

head(gap_2007)

## # A tibble: 6 x 2
##   continent gdpPercap
##   <fct>      <dbl>
## 1 Asia      975.
## 2 Europe   5937.
## 3 Africa   6223.
## 4 Africa   4797.
## 5 Americas 12779.
## 6 Oceania  34435.

anova_model <- aov(gdpPercap ~ continent, data = gap_2007)
summary(anova_model)

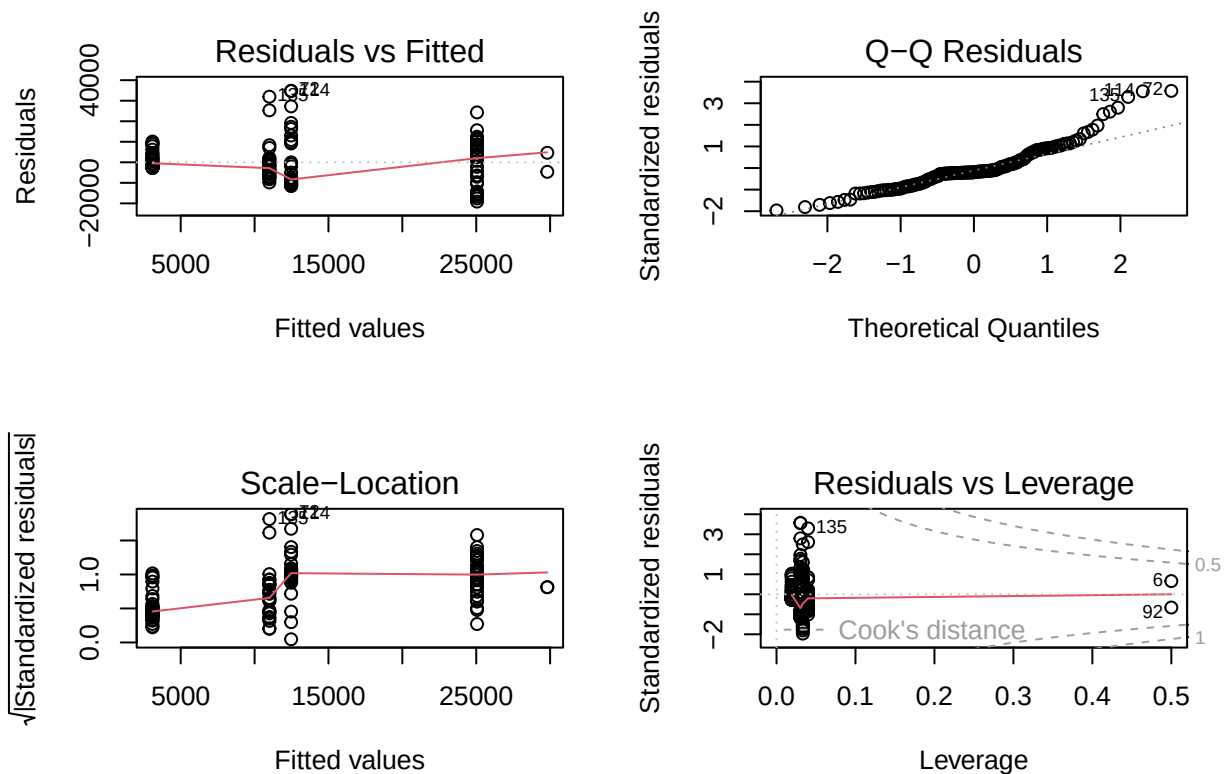
##           Df    Sum Sq  Mean Sq F value    Pr(>F)

```

```
## continent      4 9.894e+09 2.473e+09 25.24 1.13e-15 ***
## Residuals    137 1.342e+10 9.799e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# This will show 4 plots: residuals, QQ plot, etc.
```

```
par(mfrow = c(2, 2))
plot(anova_model)
```



```
library(car)
leveneTest(gdpPerCap ~ continent, data = gap_2007)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  4  6.9032 4.307e-05 ***
##      137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tukey_result <- TukeyHSD(anova_model)
print(tukey_result)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = gdpPerCap ~ continent, data = gap_2007)
##
## $continent
##      diff      lwr      upr    p adj
## Americas-Africa 7913.999 1254.048 14573.95 0.0111618
## Asia-Africa     9383.994 3293.561 15474.43 0.0003607
```

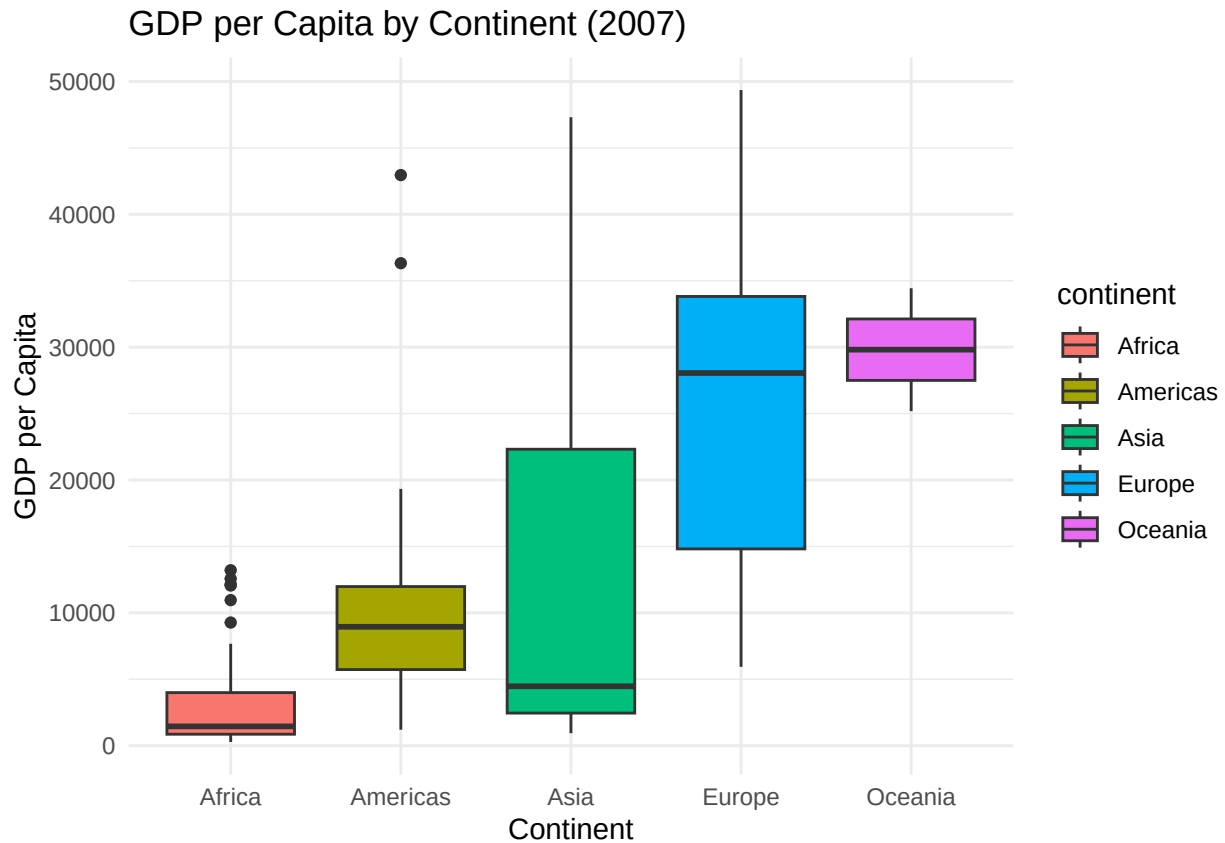
```
## Europe-Africa      21965.449  15691.486 28239.41 0.0000000
## Oceania-Africa     26721.156   7002.488 46439.82 0.0024141
## Asia-Americas      1469.995  -5785.789  8725.78 0.9805078
## Europe-Americas    14051.450   6640.940 21461.96 0.0000058
## Oceania-Americas   18807.157  -1302.018 38916.33 0.0787121
## Europe-Asia        12581.455   5678.261 19484.65 0.0000143
## Oceania-Asia       17337.161  -2590.641 37264.96 0.1201389
## Oceania-Europe     4755.707 -15228.952 24740.36 0.9648977
```

```
library(broom)
tidy(tukey_result)
```

```
## # A tibble: 10 x 7
##   term      contrast      null.value estimate conf.low conf.high adj.p.value
##   <chr>    <chr>          <dbl>      <dbl>    <dbl>    <dbl>    <dbl>
## 1 continent Americas-Africa      0      7914.    1254.    14574.    1.12e- 2
## 2 continent Asia-Africa          0      9384.    3294.    15474.    3.61e- 4
## 3 continent Europe-Africa        0  21965.   15691.   28239.    5.68e-14
## 4 continent Oceania-Africa       0  26721.    7002.   46440.    2.41e- 3
## 5 continent Asia-Americas        0   1470.   -5786.    8726.    9.81e- 1
## 6 continent Europe-Americas      0  14051.    6641.   21462.    5.79e- 6
## 7 continent Oceania-Americas     0  18807.   -1302.   38916.    7.87e- 2
## 8 continent Europe-Asia          0  12581.    5678.   19485.    1.43e- 5
## 9 continent Oceania-Asia         0  17337.   -2591.   37265.    1.20e- 1
## 10 continent Oceania-Europe      0   4756.  -15229.   24740.    9.65e- 1
```

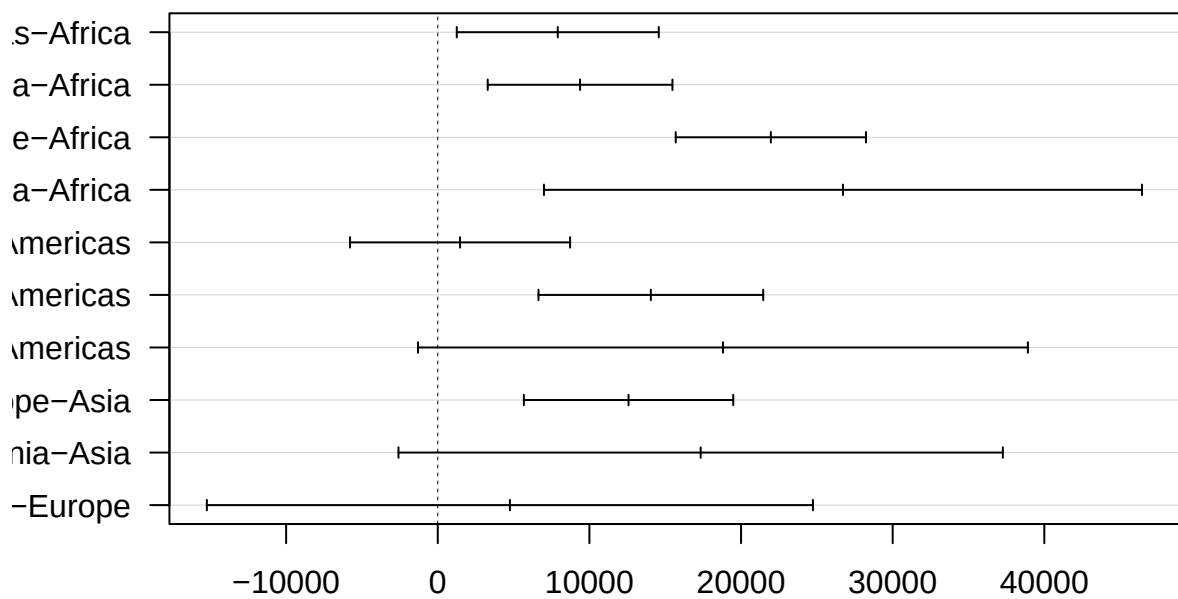
```
library(ggplot2)
```

```
ggplot(gap_2007, aes(x = continent, y = gdpPercap, fill = continent)) +
  geom_boxplot() +
  labs(title = "GDP per Capita by Continent (2007)",
       x = "Continent", y = "GDP per Capita") +
  theme_minimal()
```



```
plot(tukey_result, las = 1)
```

95% family-wise confidence level



Residuals vs Fitted Plot This plot checks for homoscedasticity and linearity. The residuals should scatter

randomly around zero; if you see patterns e.g. a funnel, you may have unequal variance.

Q-Q Plot This plot checks for normality of the residuals. Points should lie close to the reference line; if the points depart from the line, you have non-normally distributed residuals.

Scale-Location Plot This plot is also checking for homogeneity of variance. For an appropriate model, the points should scatter evenly; if you see patterns, you may not have constant variance.

Residuals vs Leverage Plot This plot identifies influential data points leveraging high leverage and large residuals. Points that lie outside Cook's distance contours should be reviewed.

Boxplot of GDP per Capita by Continent This shows the distribution (IQR), median, spread, and outliers of GDP per capita by continent for comparison.

Tukey HSD Plot Here we are showing our confidence intervals for the differences of means between continents. If the confidence intervals do not cross zero, there is a significant difference in the mean GDP for the continents compared.