

## Instructions:

- **Read** carefully the instructions for each problem. You should follow them exactly.
- **Packages:** All problems should be completed using
  - Base packages (see the list below)

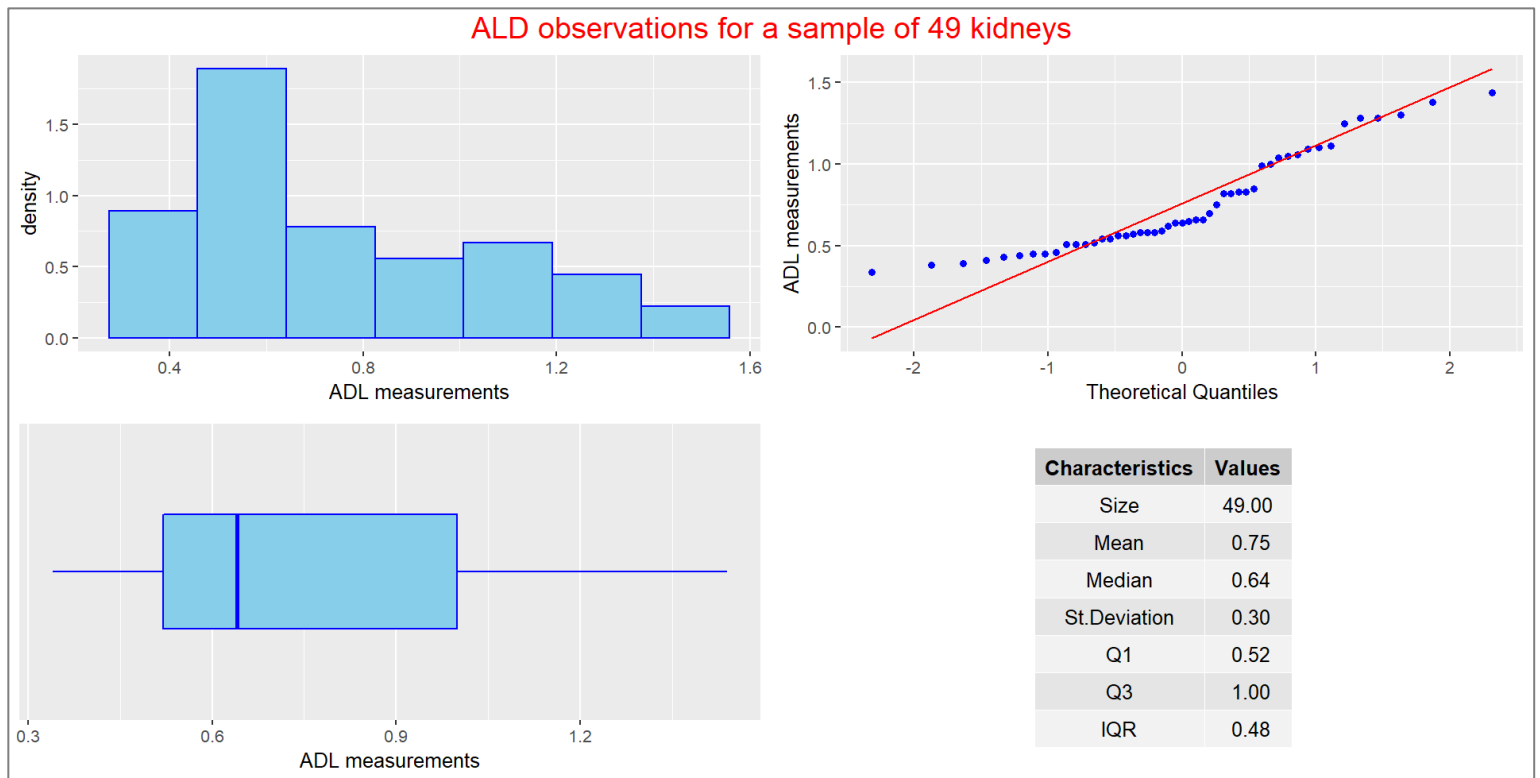
```
> search()
[1] ".GlobalEnv"          "tools:rstudio"      "package:stats"
[4] "package:graphics"    "package:grDevices"  "package:utils"
[7] "package:datasets"    "package:methods"    "AutoLoads"
[10] "package:base"
```

- ggplot2, ggpubr, readxl, writexl.
  - No other packages should be used to complete the exam.
- **Submit** a file that contains your script.
    - The working directory should contain the data file(s) used in the exam, so the files can be accessed by their names. No long paths that do not exist on my computer.
    - Load the packages and do not include install.package statement.
    - In your script, indicate the problem and part numbers. For examples, 1 a.
  - **Review** your script carefully to make sure that it does not produce error messages. I will not be correcting any errors.
1. (25 points, 5 points each part) Use ggplot2 and ggpubr to complete the following task. Automatic identification of the boundaries of significant structures within a medical image is an area of ongoing research. The paper “**Automatic Segmentation of Medical Images Using Image Registration: Diagnostic and Simulation Applications**” (*J. of Medical Engr. and Tech.*, 2005: 53–63) discussed a new technique for such identification. A measure of the accuracy of the automatic region is the average linear displacement (ALD). The paper gave the following ALD observations for a sample of 49 kidneys (units of pixel dimensions). The observations can be found in the file **ALD.txt**.

- a. Write r-code that calculates the following characteristics of the dataset ADL: the size of the data set, sample mean, sample median, sample standard deviation, the quartiles Q1 (the 25% quantile) and Q3 (the 75<sup>th</sup> quantile), and the interquartile range (also known as forth spread). Use ggpubr to create the table. Important: Do not just copy the numbers provided in the table, write your code that produces the numbers.

Characteristics	Values
Size	49.00
Mean	0.75
Median	0.64
St.Deviation	0.30
Q1	0.52
Q3	1.00
IQR	0.48

- Construct a boxplot for the data and comment on any interesting features such as the shape of distribution, presence of outliers.
- Draw a density histogram with 7 (seven) classes. Is your assumption of the shape of the distribution made in part b confirmed by the density plot? Comment on that.
- Make a normal probability plot for the data. Is it plausible that ALD is at least approximately normally distributed?
- Combine all plots and the table in one plotting area. Your final picture should look like as follows:



- (25 points, 5 points for each subplot and 5 for the combined plot) The dataset **cars** (in the package datasets) contains the speed of cars and the distances taken to stop in the 1920s. Using base packages (not ggplot2), write r-script that creates the following plot. Pay attention to the details such as the number of classes for the histograms, the axes, the labels and annotations.

# Speeds and Distances of Cars in 1920

