1 Problem 1

Before conducting any statistical analysis, it is routine for statisticians to do some **exploratory data analysis (EDA)**

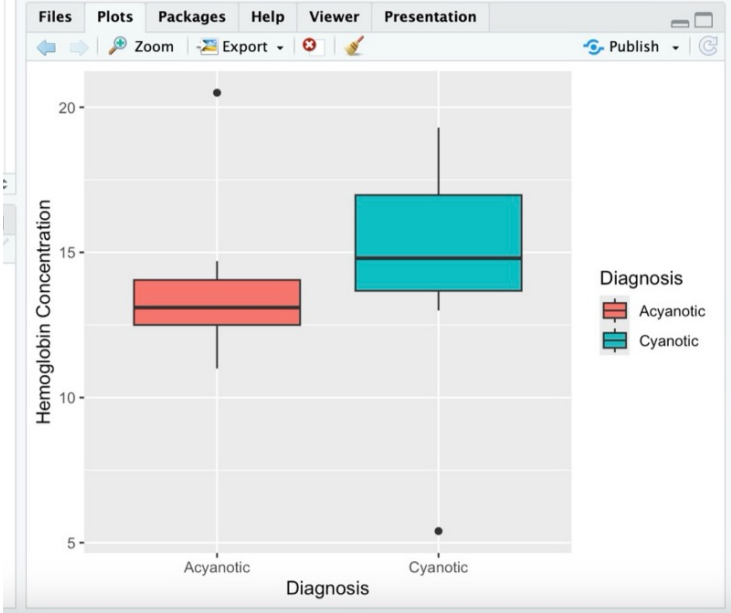
Problem 1a: Inspect the cyanotic and acyanotic data in R and fill out the table with summary statistics.

Statistic	Cyanotic	Acyanotic
Count (n)	12	19
Mean	14.65	13.45263
St. Dev.	3.501818	1.978037
Median	14.8	13.1
25%ile	13.675	12.5
75%ile	16.975	14.05
IQR	3.3	1.55
min	5.4	11
max	19.3	20.5
range	5.4~19.3	11~20.5

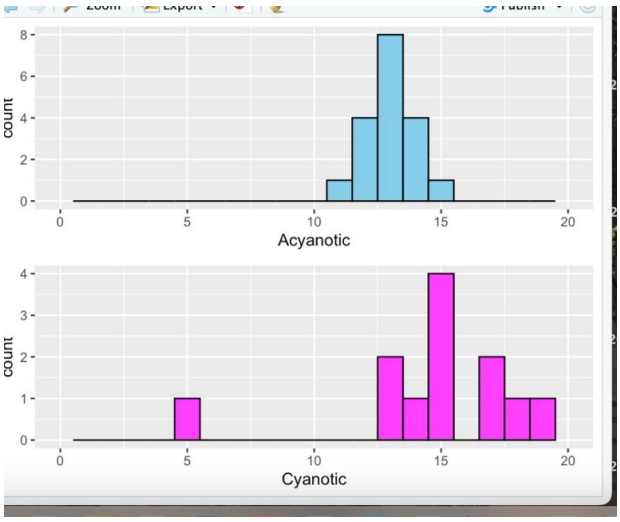
2 Problem 2

In addition to tables of summary statistics, it is also common to use graphs to conduct EDA.

Problem 2a: Create boxplots by disease subgroups for the variable hemoglobin.



Problem 2b: Create histograms by disease subgroups for the variable hemoglobin.



Problem 2c: According to these graphs, does the distribution of hemoglobin appear to differ between acyanotic and cyanotic children with congenital heart disease?

Yes. The acyanotic group is more normally distributed, while the cyanotic group is right-skewed.

3 Problem 3

We now investigate when to use arithmetic or geometric mean.

Problem 3a: Printout with values of the geometric mean by disease subgroup for hemoglobin levels. Verify the relationship that the arithmetic mean $\geq$ geometric mean within each disease subgroup.

Geometric mean is 13.33661, while as the arithmetic mean is 13.45263

Thus, the arithmetic mean > geometric mean

Problem 3b: Among the three measures of central tendency (arithmetic mean, geometric mean, and median), which best represents the central location of the data for children with acyanotic congenital heart disease? With cyanotic heart disease? Why?

According to the box plot, for the acyanotic group, the arithmetic mean is closer to the central location of the plot.(because the data is in normal distribution)
For the Cyanotic group, the median is closer to the central location of the graph.(because the data is right-skewed.)

4 Problem 4

To successfully complete this competency, answer the following review question from Triola (2006):

The clinical trial of a polio vaccine involved 200, 000 children treated with the Salk vaccine and 200, 000 other children given a placebo. Only 33 of the children in the Salk treatment group later developed polio, but 115 of the children in the placebo group later developed polio.

(a) Are the given values discrete or continuous? Discrete

(b) Identify the level of measurement (nominal, ordinal, interval, ratio) for the numbers 33 and 115. Ratio

(c) If 50 children from the treatment group are randomly selected and their average (mean) age is calculated, is the result a statistic or a parameter? **Statistics**

(d) Identify the nature (observation/experimental) and timing of the study (retrospective/prospective). What impression do you have of the relationship between vaccine and polio (causative/associative)?

experimental;
retrospective;
associative