# Exercise 1.4.2:

Select another simple random sample of size 10 from the population represented in Table 1.4.1. Compare the subjects in this sample with those in the sample drawn in Exercise 1.4.1.

Table ‑: Sample 01

|  |  |
| --- | --- |
| Subject | Age |
| 151 | 50 |
| 76 | 59 |
| 86 | 59 |
| 178 | 54 |
| 98 | 50 |
| 55 | 52 |
| 29 | 37 |
| 114 | 66 |
| 108 | 68 |
| 45 | 45 |

Table ‑: Sample 02

|  |  |
| --- | --- |
| Subject | Age |
| 149 | 53 |
| 28 | 38 |
| 144 | 43 |
| 62 | 54 |
| 88 | 57 |
| 186 | 71 |
| 4 | 44 |
| 172 | 58 |
| 177 | 61 |
| 123 | 68 |

1. Are there any subjects who showed up in both samples?

No.

1. How many?

0.

1. Compare the ages of the subjects in the two samples.

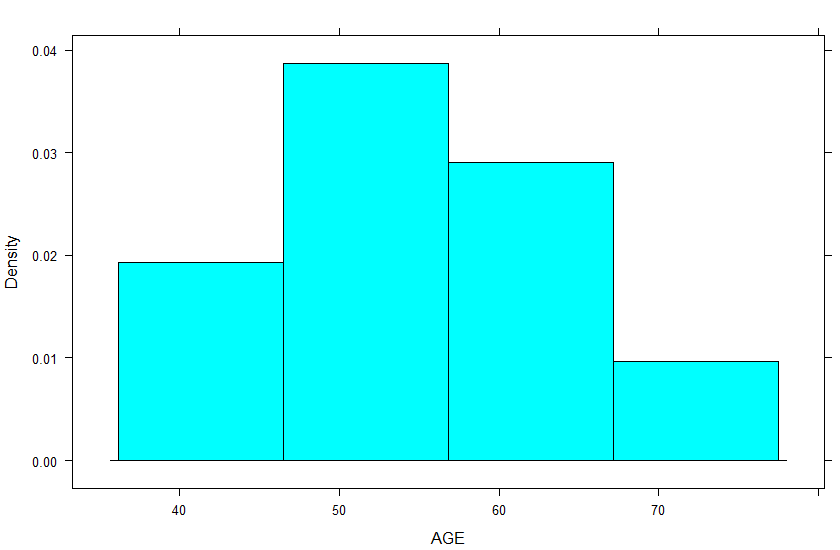


Figure .: Sample 01 histogram

Sample 01 has a mean of 54 and a standard deviation of 9.404.

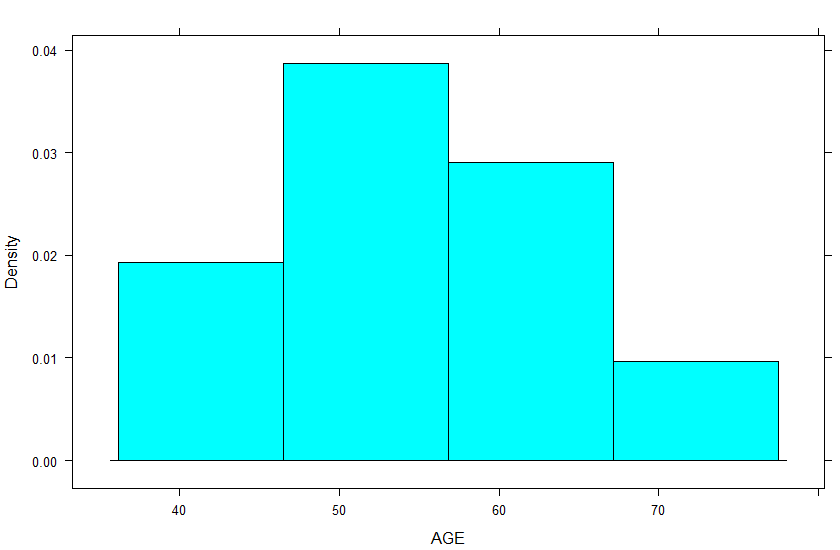


Figure .: Sample 02 histogram

Sample 02 has a mean of 55.5 and a standard deviation of 10.709.

1. How many ages in the first sample were duplicated in the second sample?

Two ages were duplicated. The ages 54 and 68 appeared in both samples.

# Exercise 1.4.4:

Construct an example where it would be appropriate to use stratified sampling. Discuss how you would use stratified random sampling and stratified sampling proportional to size with this example. Which do you think would best represent the population that you described in your example? why?

Compare the income of people in two different geographical regions. Stratify the data by education level. Stratification proportional to size would be appropriate in this example as it is likely that the populations with higher levels of education are successively smaller, and if a sample of the same size was taken from each strata, higher educated groups would be over-represented.

# Chapter 1 exercise 6

For each of the following variables, indicate whether it is quantitative or qualitative and specify the measurement scale that is employed when taking measurements on each:

1. Class standing of the members of this class relative to each other

Qualitative - Ordinal scale

1. Admitting diagnosis of patients admitted to a mental health clinic

Qualitative - Nominal scale

1. Weights of babies born in a hospital during a year

Quantitative - Ratio scale

1. Gender of babies born in a hospital during a year

Qualitative - Nominal scale

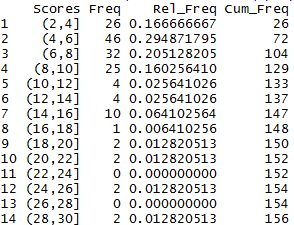
1. Range of motion of elbow joint of students enrolled in a university health sciences curriculum

Quantitative - Ratio scale

1. Under-arm temperature of day-old infants born in a hospital

Quantitative - Interval scale

# Exercise 2.3.2

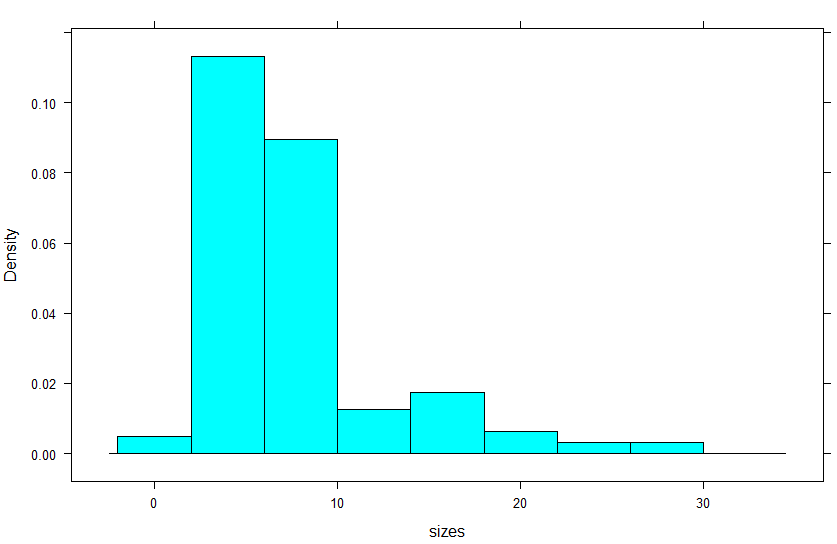


.: Frequency distribution, relative frequency distribution, cumulative relative frequency distribution

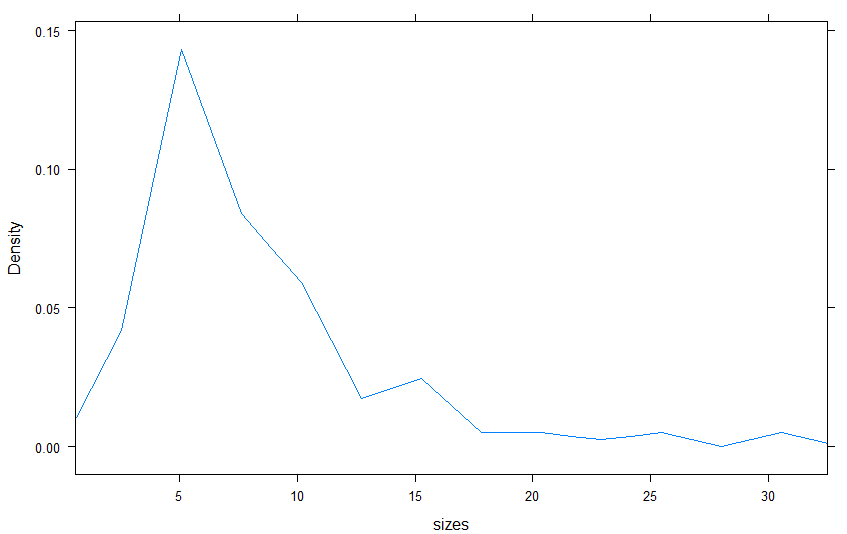
How many observations are less than 20? 150.

What proportion of the measurements are greater than or equal to 25? 2.56%

What percentage of the measurements are either less than 10.0 or greater than 19.95? 137



.: Histogram



.: Frequency polygon

# Exercise 2.5.6

Mean: 41.5

Median: 39.5

Mode: 19

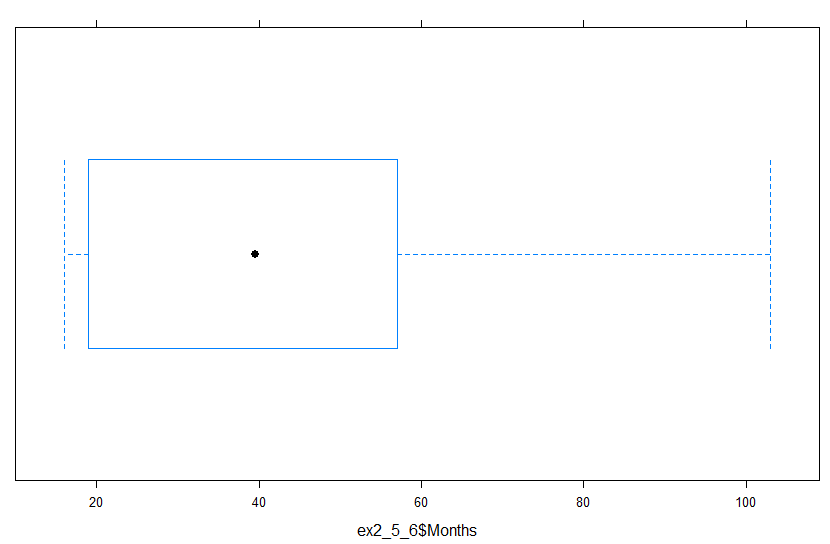
Range: 87

Variance: 490.26

Standard Deviation: 22.14

Coefficient of variation: 0.533

Interquartile range: 36.5



.: Box and whisker plot

The measure of central tendency that would be most appropriate for this data set would be the median. This is because the data is clustered near the low end and extreme, uncommon values that are very high skew the mean upwards. The mean would give an estimate of the center of the data set that is too high.

# Review exercise 12 (hand entry)

## Control

Mean: 126.1

Median: 123

Standard Deviation: 21.82

Variance: 476.10

## SCI

Mean: 133.1

Median: 130

Standard Deviation: 32.18

Variance: 1035.43

# Review exercise 26

Indicate for the following variables which youy think would be a better measure of central tendency, the mean, the median, or mode. Justify your choice.

## Annual incomes of licensed practical nurses in the Southeast.

Median. This is a case where extreme values could skew the measure of center.

## Diagnoses of patients seen in the emergency department of a large city hospital.

Mode. The mode is uniquely able to measure the central tendency of non-numerical data.

## Weights of high-school male basketball players

Mean. Pysiology limits extreme values and the weights of players is likely continuous data and somewhat normally distributed. A good fit for a mean.

# North Carolina Birth Records

## Mother's age

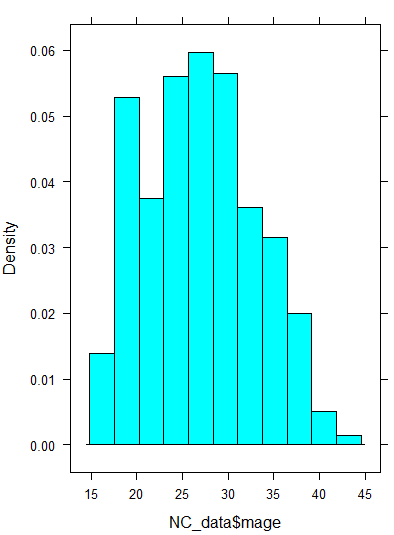
Mean: 26.915

Median: 26

Standard deviation: 6.107

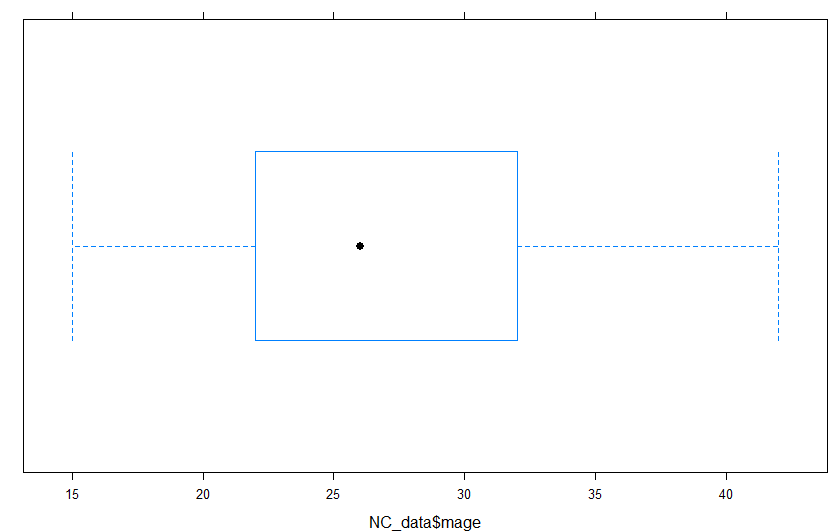
IQR: 10

Range: 27



.: Histogram of maternal age at time of child's birth

The distribution of maternal age looks reasonably normal perhaps with a slightly higher occurance of younger mothers.



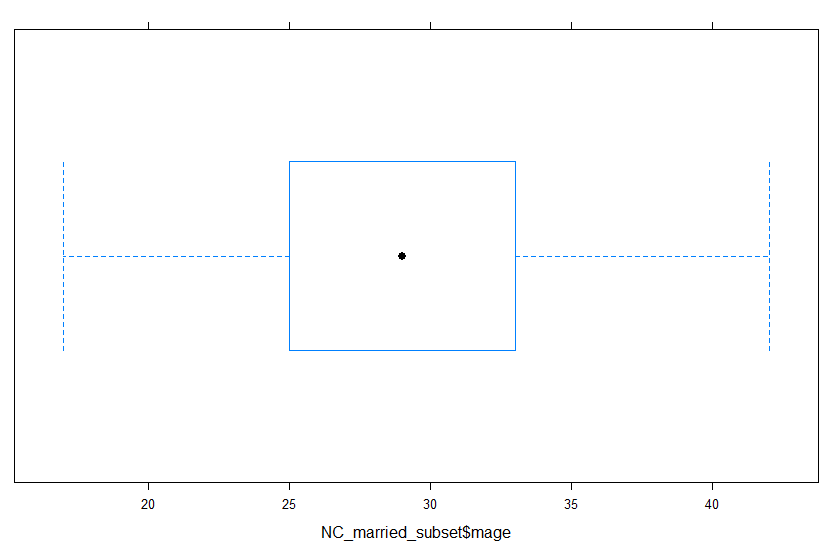
.: Box and whisker plot of maternal age at the time of child's birth

Skewness: 0.2107

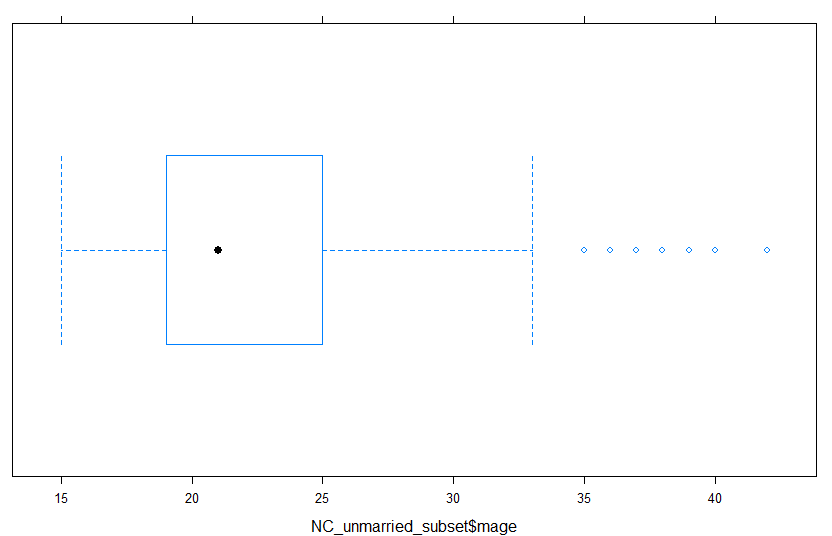
This positive skew indicates that permutations of lower age are somewhat more common than those of higher ages. Maternal age is clustered toward the lower end of the distribution. This matches the observation based on the the histogram.

Kurtosis: 2.203

This kurtosis means that the distribution of MAGE is slightly platykurtic which is to say that it has a flattened appearance when compared to a standard normal distribution. this suggests that extreme ages are more common than a normal distribution would suggest.



.: Box and whisker plot of married maternal age



.: Box and whisker plot of unmarried maternal age

For this variable only, construct side-by-side BW plots for women who are and are not married. Do you see a difference in ages in the two groups? Which group has more variability? Are the results surprising?

The unmarried mothers are younger on average, but that data subset also shows more variability. Maternal age range appears to be the same for both groups. This is not surprising as the range is likely biologically determined, and the probability that a mother is married increases with age.

## Weeks of gestation

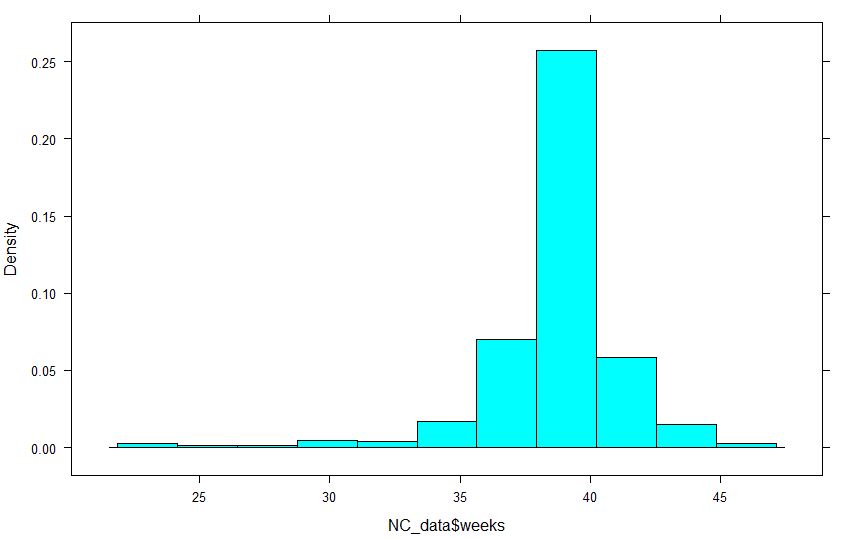
Mean: 41.5

Median: 39

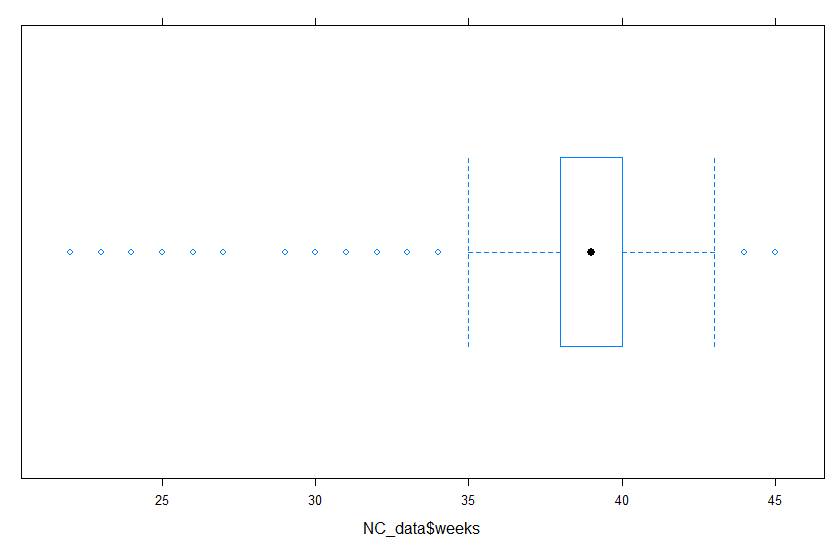
Standard deviation: 2.716

IQR: 2

Range: 23



.: Weeks gestation histogram



.: Weeks gestation box and whisker plot

Skewness: -2.090

This data set is skewed very much to the left. This indicates that weeks of gestation is clustered toward the higher end of the distribution and that very low extremes occur while very high extremes do not.

Kurtosis: 12.231

This very high kurtosis indicates that the data are clustered very near the mean and that the range is very large with respect to the spread of this cluster. The distribution is very "peaky."

## Weight gained

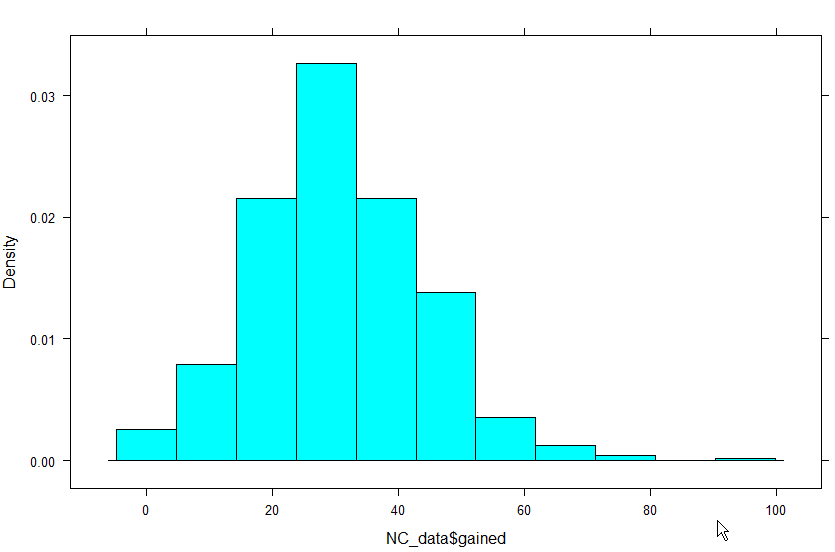
Mean: 30.58

Median: 30

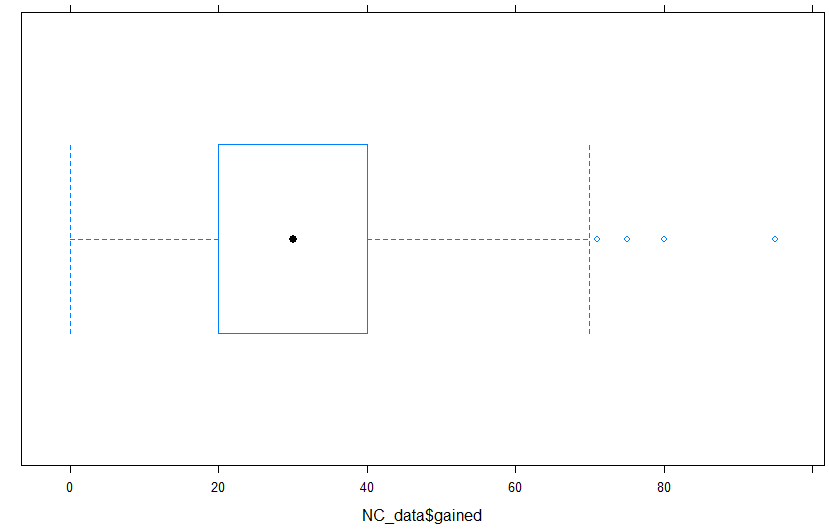
Standard deviation: 13.649

IQR: 10

Range: 95



.: Maternal weight gain histogram



.: Box and whisker plot of maternal weight gain

Skewness: 0.419

The data are skewed to the right which makes sense as the amount of weight gain is practically infinite in the positive direction while losing weight would be relatively rare resulting in a cluster on the lower end of the distribution.

Kurtosis: 3.802

This data closely matches a standard normal distribution

## Weight of child (ounces)

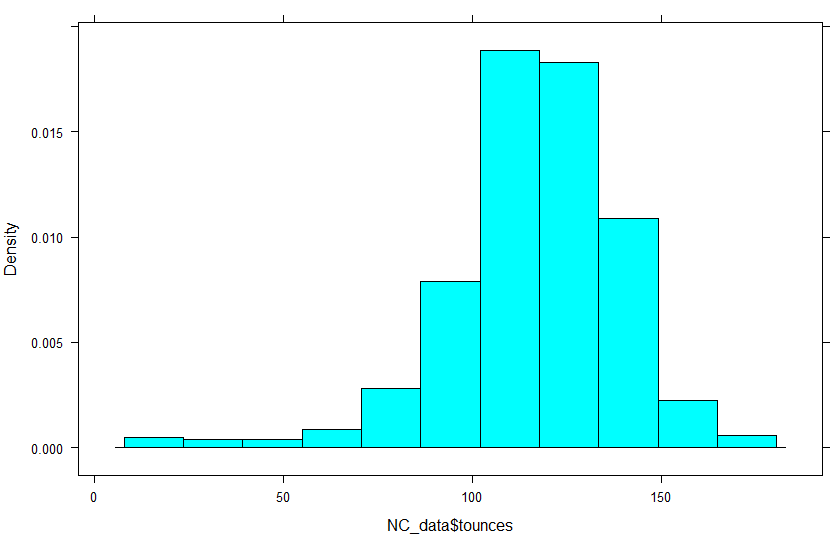
Mean: 116.38

Median: 118

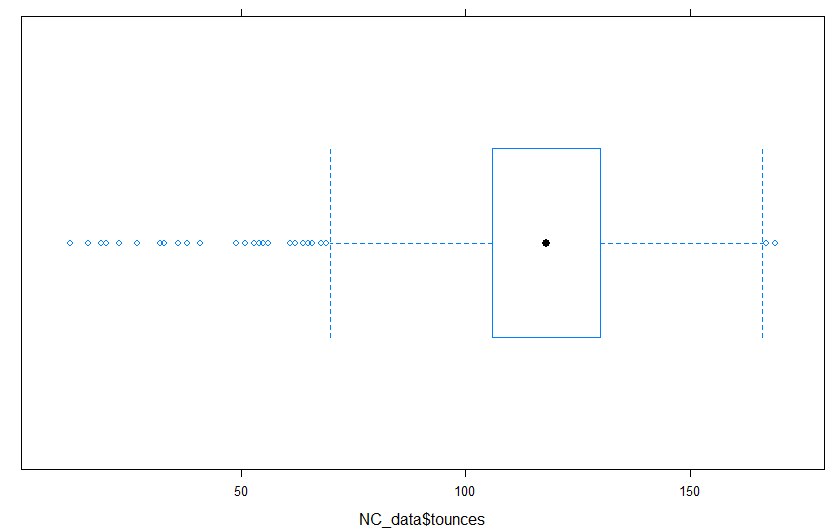
Standard deviation: 22.539

IQR: 24

Range: 157



.: Birth weight (ounces) histogram



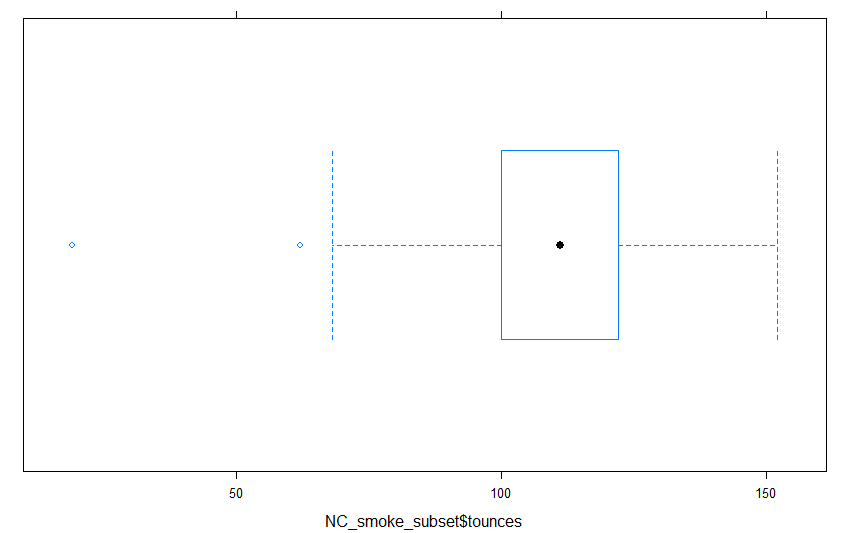
.: Birth weight (ounces) box and whisker plot

Skewness: -1.051

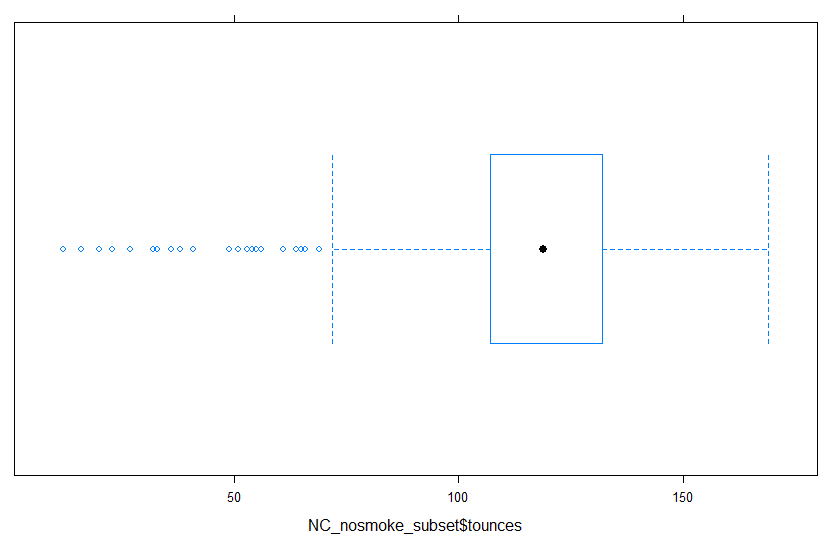
These data are skewed to the left. This makes sense as there is a "maximum" birth weight that is practical.

Kurtosis: 6.158

The data is clustered around the mean with a relatively small amount of variation.



.: Birth weight prenatal smoking



.: Birth weight no prenatal smoking

Surprisingly, there didn't appear to be a large difference in average birth weight. However, there was a lot more variability, particularly favoring low birth weight for moms who smoked during pregnancy.

## Weight of child (grams)

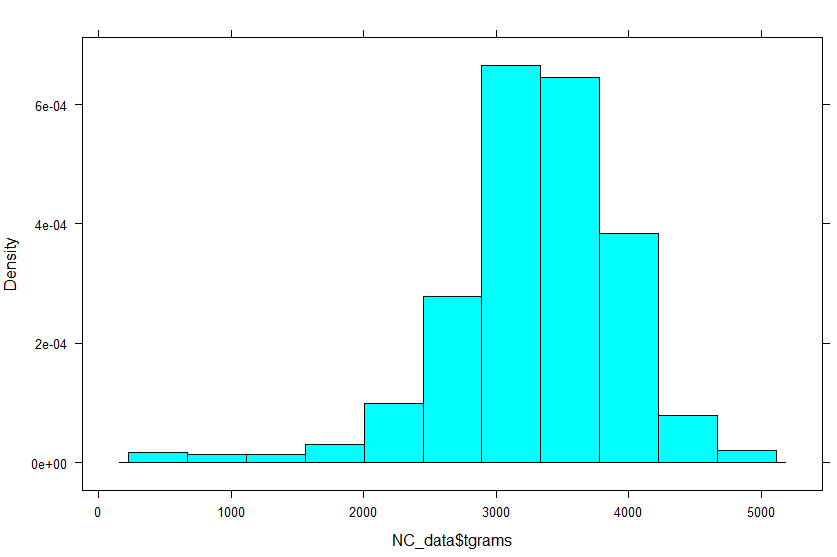
Mean: 3299

Median: 3345

Standard deviation: 638.97

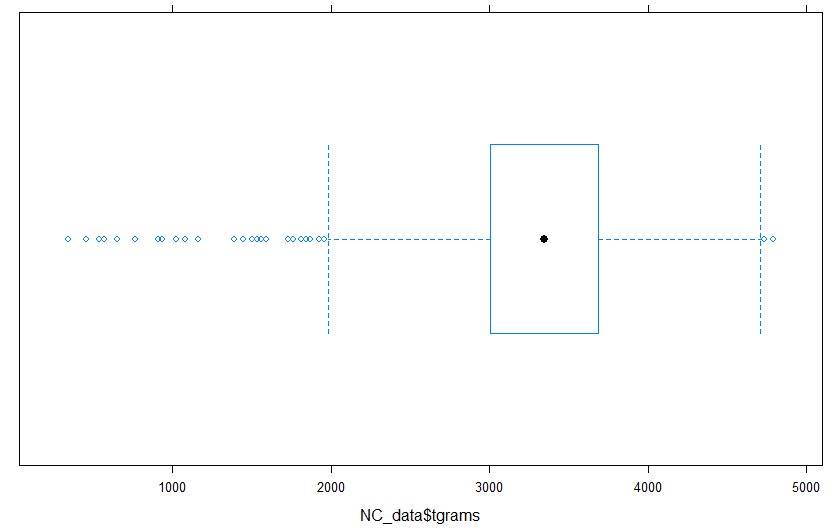
IQR: 24

Range: 4451



.: Birth mass (grams) histogram

Yes, this looks the same as Tounces. Same data, different units.



.: Birth mass (grams) box and whisker plot

Skewness: -1.051

Kurtosis: 6.158

Both of these measures are exactly the same as the previous data.