Note: you will get a R code cheat sheet that includes the functions for the exam.

Imagine you are a consultant for a record company, and they have no idea what statistics are, but know that they need them. Given the following data, answer their questions with *R* output or in jargon free lingo. You can assume they know what a mean score consists of, but not necessarily things like variance or standard deviation.

We’ve recently collected data from our artists about their record sales, and weeks on the charts (see table below). We wanted some pretty pictures and descriptions of the data to be able to put into our presentation for the CEO in a couple weeks. Record sales are in the thousands.

|  |  |  |
| --- | --- | --- |
| Type of Artist | Record Sales | Chart Weeks |
| Pop | 105 | 5 |
| Country | 500 | 10 |
| Pop | 75 | 3 |
| Pop | 203 | 5 |
| Pop | 100 | 5 |
| Country | 109 | 5 |
| Country | 650 | 15 |
| Country | 400 | 4 |
| Pop | 65 | 3 |

For the first column only:

* What is the variable?
* What are the levels?

For the rest of the columns:

* What type of variable is record sales using the NOIR system?
* How might the company operationally define chart weeks?

Create a histogram of record sales.

* Do there appear to be outliers?
* What shape is the histogram?
* Does the distribution have any skew?

Create a frequency table of chart weeks.

What are the mean, median, mode(s), and interquartile range for record sales and chart weeks?

What is the *biased* standard deviation and variance estimate for chart weeks?

How does that differ than the *unbiased* estimates (you can describe rather than give numbers)?

Create a graph of record sales and chart weeks with a line of best fit.

Create a graph for the different types of artists’ average record sales. Who seems to sell more records?