

CSC 315, Exam I

Exam I Notes

- You may bring one page of notes (front and back) to the exam. This page may be handwritten or typed.
- Computer access will not be permitted during the exam.
- Cell phones must be put away at all times
- Don't hesitate to contact me if you have any questions!

Exam I Concepts

- R coding
 - Creating vectors (*c*, *seq*, *a:b*, *rep*)
 - Creating lists
 - Selecting or changing elements:
 - That are values of a vector
 - That are rows, columns, or values of a matrix or data frame
 - That are elements of a list
 - Viewing or changing names:
 - Names of a list
 - Row names of a matrix or data frame
 - Column names of a matrix or data frame
 - Data frames:
 - Filtering and selecting columns using *dplyr*
 - The *split* function
 - Writing user-defined functions
 - Using *apply*, *lapply*, and *sapply*
- Graphical and Numerical Summaries
 - Generating a frequency and relative frequency tables from raw data
 - Constructing bar graphs (including Pareto charts) from raw data and from frequency or relative frequency tables
 - Interpreting histograms (unimodal vs. bimodal; symmetric vs left-skewed vs. right-skewed)
 - Finding the following statistics in R, and interpreting them
 - mean, median, variance, standard deviation, minimum, maximum, percentiles
 - Understanding properties of mean vs median
 - Constructing side-by-side boxplots and interpreting them
- Associations
 - Constructing contingency tables and conditional proportions from raw data
 - Constructing stacked and side-by-side bar graphs from raw data or from contingency tables or conditional proportions
 - Constructing a scatterplot, including with the regression line
 - Finding the correlation
 - Linear models:
 - Fitting the linear model
 - Interpreting the y-intercept and slope
 - Making predictions, and understanding extrapolations

- ggplot:
 - the layers *geom_bar*, *geom_col*, *geom_boxplot*, *geom_point*, *geom_smooth*,
 - adding a title and x- and y-axis labels
 - coloring points or bars based on the data (aesthetics), or based on a set of values
 - changing the theme
 - hiding the legend
 - Using and understanding *facet_grid*

Example questions (also see the Practice R script):

1) Display the 3rd element from a vector named *last_name*.

For questions (2) and (3), consider a data frame named *students*, whose first 4 rows are shown below:

```
> head(students)
# A tibble: 6 × 7
  Gender `College GPA` `HS GPA`   FB `Marijuana-Legalization` Same-Sex...1 Alcohol
  <chr>      <dbl>    <dbl> <dbl> <chr>                <chr>      <dbl>
1 Female      3.3      3.5    1 Agree                Agree       4
2 Male        2.5      2.25   2 Agree                Agree       2
3 Male        2.5      3       1 Agree                Agree       7
4 Male        3.12     2.3    2 Agree                Disagree    3.5
```

2) Find the College GPA of the 10th individual

3) Find the median HS GPA (assume that missing values are possible)