

Post Course Recap: Introduction to R

(With Tidyverse)

Try to perform the operations below using the data in the Introduction to R data folder.

- Open a new script file (and restart RStudio if you still had it running from before)
- Set your working directory to be the directory with the R intro data in it
- Import the functions from the tidyverse package into your script

Trumpton

- Load in the “`trumpton.txt`” tab delimited file into a tibble and save it into a variable
- Plot out a scatterplot of the `Age` vs `Weight`, what generally happens as you get older?
- There is only one individual who weighs more than 100kg, make a selection to find this person and then just show their first and last names
- Plot a barplot (using `geom_col`) where the x aesthetic is the last name, and the y aesthetic is the Age. Make it so the bars are filled with `magenta2` and the lines are coloured `black`

Child Variants

- Load the child variants dataset
- Select all of the rows (variants) which occur in the first 5Mbp of Chr X
- From this set of variants plot out a scatterplot of `MutantReads` vs `Coverage` and colour it by quality. What do you notice about the poor quality calls?
- Filter the variants on chr 1 which have a valid dbSNP id. Plot out their position vs their coverage as a line graph. Colour the line grey and make it 1 unit thick.
- Repeat the last plot, but remove any variants with a coverage of over 200