

Pipes quiz

CMSC498 Fall 2016

September 14, 2016

Name(s):

UID(s):

You are working for real estate website Zillow and you want to calculate income distributions in a geographically sensible way. You have a `data.frame` with two columns `income` and `address` where addresses are in form “Address line 1\nAddress line 2\nCity, State, Zip” for some number of households. You think that calculating income based on the first three digits of the zip code makes sense and you want to write a pipeline to extract that information.

You have access the following functions:

- `select(data_frame, columns)`: takes a data frame and returns a data frame with only the given `columns`
- `split_address(addresses)`: takes a vector of addresses in format above and returns a `data.frame` with columns `line_1`, `line_2`, `city`, `state`, `zip`
- `extract_prefix(prefix_size, vector)`: takes a string vector `vector` and returns the prefix of size `prefix_size` from each string in the vector

Write a short program using the pipe `%>%` operator that generates a `data.frame` containing the first three digits of the zip code and income for each household in the dataset.