

# ST 538 Project 1

Group 3

2024-04-19

```
temp <- tempfile()
download.file("https://www2.census.gov/programs-surveys/acs/data/pums/2022/5-Year/csv_por.
data <- read_csv(unz(temp, "psam_p41.csv"))
```

Rows: 205072 Columns: 290

-- Column specification -----

Delimiter: ","

chr (27): RT, SERIALNO, SPORDER, PUMA10, PUMA20, JWTRNS, SCHG, SCHL, ANC1P,...

dbl (263): DIVISION, REGION, ST, ADJINC, PWGTP, AGE, CIT, CITWP, COW, DDRS,...

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
unlink(temp)
```

We are interested in studying the relationships between a teachers salary and other attributes in the data set. We will start off with a large number of variables and attempt to do variable selection methods to determine which variables are the most important in determining a teachers salary.

Variable Name	Description
WAGP	Wages or salary past 12 months
SEX	Sex of person(m/f)
AGEP	Age of person
SCHL	Educational attainment
SCIENGP	Field of degree(na-less than bachelor, 1 science/engineering, 0 not)