STA 210 - Final Project

Zakk Heile & Julia Healey-Parera

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4      v readr     2.1.5
v forcats   1.0.0      v stringr   1.5.1
v ggplot2   3.4.4      v tibble    3.2.1
v lubridate 1.9.3      v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becon
```

```r
library(broom)
library(glm2)
library(dplyr)
library(Stat2Data)
library(pROC)
```

```
Warning: package 'pROC' was built under R version 4.3.3


Type 'citation("pROC")' for a citation.

Attaching package: 'pROC'

The following objects are masked from 'package:stats':

    cov, smooth, var
```

```r
library(yardstick)
```

Warning: package 'yardstick' was built under R version 4.3.3

Attaching package: 'yardstick'

The following object is masked from 'package:readr':

    spec

```r
library(ggplot2)
library(janitor)
```

Warning: package 'janitor' was built under R version 4.3.3

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test

```r
library(here)
```

here() starts at C:/Users/think/OneDrive/Documents/STA210Final

```r
library(fs)
library(withr)

#edu <- read.csv("data/english_education.csv")
edu <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/mast
```

Rows: 1104 Columns: 31
-- Column specification ---
Delimiter: ","

```
chr (13): town11cd, town11nm, size_flag, rgn11nm, coastal, coastal_detailed,...
dbl (18): population_2011, ks4_2012_2013_counts, key_stage_2_attainment_scho...

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.
```

```r
working_dir <- here::here("data")

xls_path <- withr::local_tempfile(fileext = ".xlsx")
download.file(
  "https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/educationandchildcare/dat
  xls_path,
  mode = "wb"
)

english_education <- readxl::read_xlsx(xls_path, sheet = "Data", na = "*") |>
  janitor::clean_names()

readr::write_csv(
  english_education,
  fs::path(working_dir, "english_education.csv"))
```

```r
#transforming to log odds
edu <- edu %>%
  mutate(log_odds_ks4asy = log(key_stage_4_attainment_school_year_2012_to_2013 / (1 - key_
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `log_odds_ks4asy =
  log(key_stage_4_attainment_school_year_2012_to_2013/(1 -
  key_stage_4_attainment_school_year_2012_to_2013))`.
Caused by warning in `log()`:
! NaNs produced
```

```r
##m1 <- lm(log_odds_ks4asy ~ population_2011 + factor(income_flag), data = edu)

#summary(m1)

#transforming back
#predicted_log_odds <- predict(m1, type = "response")
#predicted_proportions <- exp(predicted_log_odds) / (1 + exp(predicted_log_odds))
```