

# Dplyr Join Lab

*Abbie M Popa*

## Set-Up

You will need the following packages

```
library(dplyr)
library(ggplot2)
```

You will need the following data sets

```
cust_ids <- read.csv("https://github.com/abbiepopa/BSDS100/raw/master/Data/cust_ids.csv")
cust_rev <- read.csv("https://github.com/abbiepopa/BSDS100/raw/master/Data/cust_rev.csv")
no_consent <- read.csv("https://github.com/abbiepopa/BSDS100/raw/master/Data/no_consent.csv")
```

The `cust_ids` data set contains a list of customers (`id_num`), whether they saw no ad, a picture ad, a text ad, or a video ad (`ad_type`) on a given visit to the site, and what location (`location`) they were viewing the website from.

The `cust_rev` data set contains a list of customers(`id_num`) and how much money they have spent, in total, through clicking on ads on the website.

The `no_consent` data set contains a list of customers who have asked to have their data removed from your training data under California's new consumer privacy law.

You want to determine what affect `ad_type` and `location` have on `money_spent`. You also must remove customers who have asked that their data be removed. Combine the data using functions we learned from `dplyr`. Then, visualise the effect of `ad_type` and `location` using plots.

Based on your plot(s), what effect do you think `ad_type` and `location` have on `money_spent`?