Obtaining data of interest

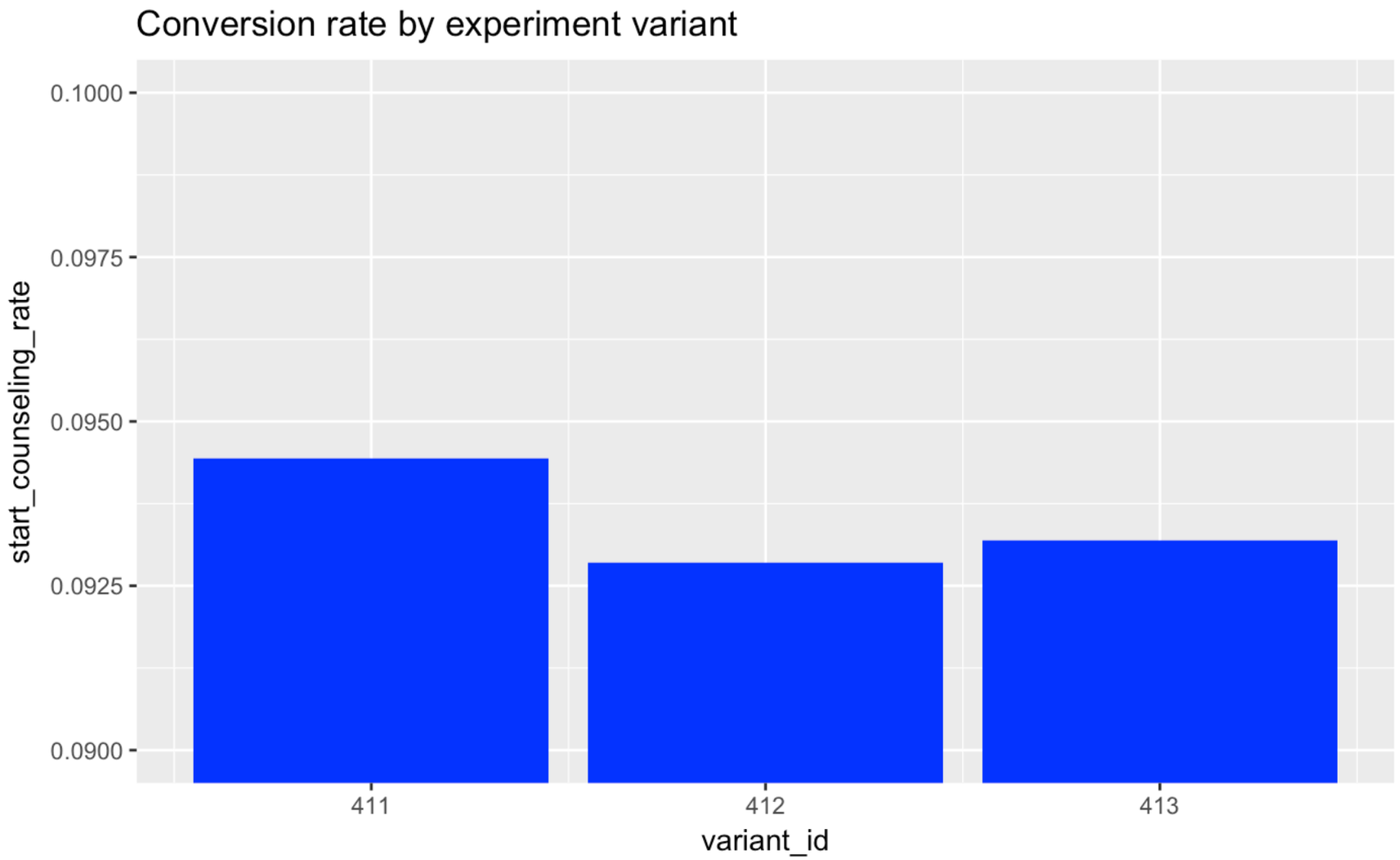
1. Joined **`account`** and **`account_variant`** tables by **`id`** (i.e., **`account_id`** in the **`account_variant`** table)

2. Selected **`id`**, **`is_mobile`**, **`country_id`**, **`entered_cc`**, **`balance`**, **`started_video`**, **`finished_video`**, **`source`**, and **`variant_id`** columns

3. Created two binary variables, **`usa`** and **`is_balance`**, that indicate whether user is from USA and whether user has non-empty balance, respectively

4. Filtered out nonsense observations where user did not start the video but somehow finished the video (albeit only 12 of these)

Problem: Who should see the video?



- At the highest level, not showing the video at all (i.e., **`variant_id`** = 411) seems best, followed by showing it at the bottom (i.e., **`variant_id`** = 413)

- However, by coupling **`variant_id`** with other categorical variables to comprise grouping variables, we can better answer who should see the video and where the video should be placed (see example on next slide)

Solution: A programmatic approach to grouping by then summarizing

Achieved: Found all values for grouping variables that, when coupled with **`variant_id`**, result in a higher conversion rate for variants 412 or 413 than for variant 411

Example: This is one of 64 combinations outputted by my code

```
[1] "Show video at the TOP when"
# A tibble: 1 x 3
  is_mobile   usa source
      <int> <dbl>  <chr>
1         0     0  other
```

Interpretation: The video should be shown at the top for users who are not on mobile, are not from the US, and who are from source "other"

```
> dat %>% group_by(is_mobile, usa, source, variant_id) %>%
+   summarize(start_counseling_rate = mean(entered_cc)) %>%
+   filter(is_mobile == 0, usa == 0, source == "other")
# A tibble: 3 x 5
# Groups:   is_mobile, usa, source [1]
  is_mobile   usa source variant_id start_counseling_rate
      <int> <dbl>  <chr>      <int>                 <dbl>
1         0     0  other        411             0.2000000
2         0     0  other        412             0.5000000
3         0     0  other        413             0.4545455
```

Note: These rates may be subject to small sample size and/or may not be statistically significant, but interpretation is still valid and follow-through implementation wouldn't hurt

Miscellaneous

Other insights

- Though the video itself does not help in getting conversions, whether the user watched the video is a good indicator of whether the user will start counseling

- That is, users who start and finish watching the video are most likely to convert, followed by users who start but don’t finish, followed by users who don’t start the video at all

Key assumptions

- Assumed the **`entered_cc`** column in the **`account`** table represents whether user started counseling

- Assumed the users involved in this experiment accurately represent all other users

- Assumed users were split into variant groups such that each group has similar proportions when looking across demographic variables