Diversity in PhD Theses

Sharon Velpula

September 13, 2022

Data

```
pi <- read_csv('https://sldr.netlify.app/data/phd_innovation.csv')
```

```
## Rows: 4195 Columns: 8
## -- Column specification --------------------------------------
## Delimiter: ","
```

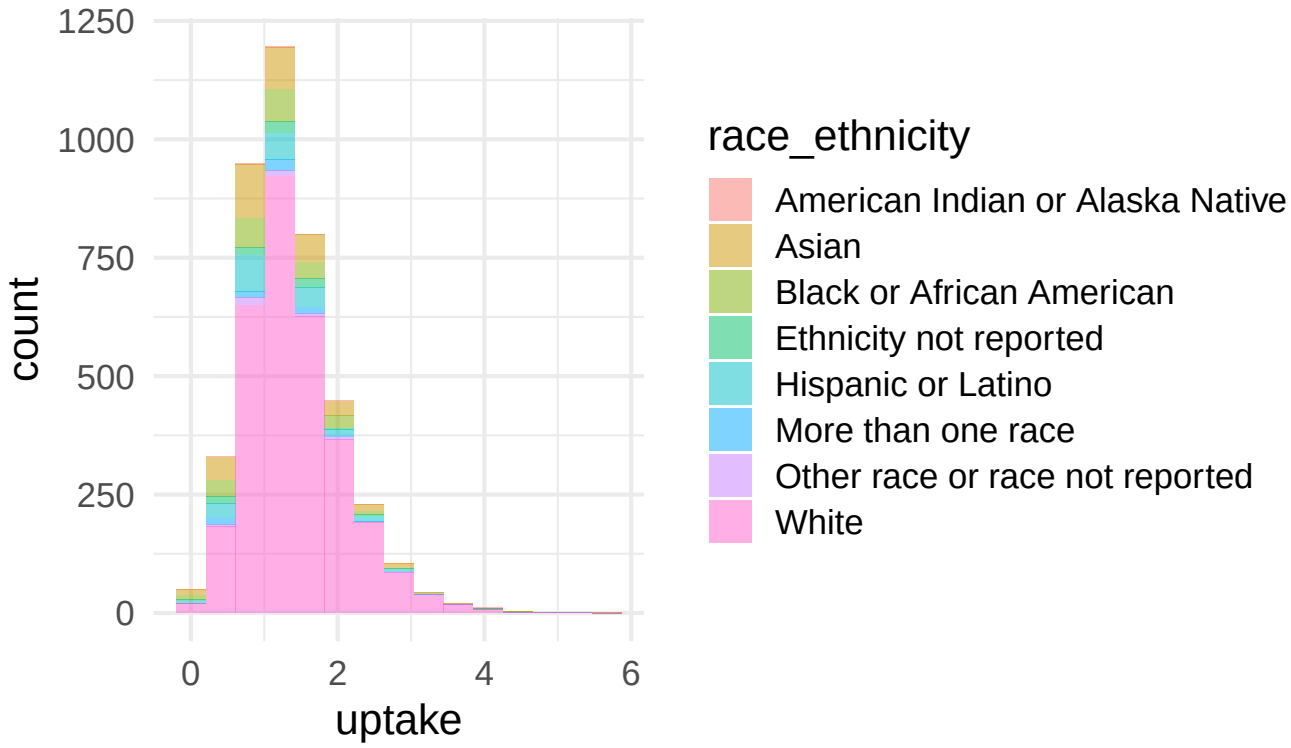
```
## chr (3): field, race_ethnicity, sex
## dbl (5): year, prop_same_race, prop_same_sex, novelty, uptake
##
## i Use `spec()` to retrieve the full column specification for this da
## i Specify the column types or set `show_col_types = FALSE` to quiet
```

```
pi
```

```
## # A tibble: 4,195 x 8
##     year field  race_ethnicity sex   prop_same_race prop_same_sex no
##    <dbl> <chr>  <chr>          <chr>          <dbl>         <dbl>
##  1  2006 Life ~ White          Fema~         0.757         0.516
##  2  2001 Mathe~ Asian          Fema~         0.118         0.234
##  3  2011 Mathe~ White          Fema~         0.735         0.249
##  4  2016 Life ~ Black or Afri~ Male          0.0586        0.449
##  5  2011 Mathe~ Black or Afri~ Male          0.0318        0.751
##  6  2001 Human~ White          Male          0.806         0.512
##  7  2016 Psych~ White          Fema~         0.710         0.587
##  8  2006 Human~ Asian          Fema~         0.0528        0.497
```

```
##  9  2001 Engin~ White                 Male              0.707              0.831
## 10  2006 Mathe~ White                 Fema~             0.731              0.253
## # ... with 4,185 more rows
```

```
gf_histogram(~ uptake, fill = ~ race_ethnicity, bins = 15, data = pi)
```



I chose a histogram because I wanted to visualize both the categorical variable race_ethnicity and the quantitative variable uptake. We can see that up till a

certain point, the increase in uptake shows an increase in the count of different ethnicity groups. However, after that point, increasing uptake shows a decrease in the count of the ethnicity groups. Just by observation, it looks like most people experience an uptake between 1 and 1.5, with the Asians experiencing it the most.