

Estimation of Doctors in Each State

Yanzun Jiang

```
# Count the number of doctors by state (STATEICP)
state_doctor_count <- data %>%
  filter(EDUCD == 116) %>%
  group_by(STATEICP) %>%
  summarise(n = n()) %>%
  arrange(desc(n))

# Record number of doctors in CA
doctor_in_ca <- state_doctor_count %>%
  filter(STATEICP == 71) %>%
  pull(n)

state_respondant_count <- data %>%
  group_by(STATEICP) %>%
  summarise(n = n()) %>%
  arrange(desc(n))

total_respondents_california <- 391171
state_doctor_count %>% mutate(
  estimated = round(total_respondents_california * n / doctor_in_ca),
  actual = state_respondant_count$n,
  diff = actual - estimated
)
```

```
# A tibble: 51 x 5
  STATEICP      n estimated actual  diff
  <dbl> <int>      <dbl> <int> <dbl>
1      71  6336   391171 391171     0
2      49  3216   198549 292919 94370
3      13  2829   174656 217799 43143
4      43  2731   168606 203891 35285
5       3  2014   124340 132605  8265
```

6	14	1620	100015	128046	28031
7	52	1608	99274	120666	21392
8	40	1531	94521	109349	14828
9	21	1457	89952	109230	19278
10	44	1451	89582	101512	11930

i 41 more rows