How about more categories?

Contrasts, the default in most lm() functions

- There are a few ways of modeling predictors with many categories:
  - **Contrasts**: Each category is compared to a **baseline**, and the coefficients are comparisons between baseline and levels
  - **One-hot**: coefficients are means of each level of the predictor
  - **Residuals**: an overall mean is measured, and coefficients are differences between each level and the global mean

```
> fit_contrasts = stan_glm(y ~ x, data = df)
> summary(fit_contrasts)[1:3, 1:3]
             mean mcse   sd
(Intercept) 0.72    0 0.13
xB          1.32    0 0.20
xC          2.20    0 0.186
```

Q: How do we get estimates for the mean in each class?

How about more categories?

One-hot

- There are a few ways of modeling predictors with many categories:
  - **Contrasts**: Each category is compared to a **baseline**, and the coefficients are comparisons between baseline and levels
  - **One-hot**: coefficients are means of each level of the predictor
  - **Residuals**: an overall mean is measured, and coefficients are differences between each level and the global mean

```
> x
[1] "B" "A" "A" "B" "C" "B" "B" "C" "B"
> onehot
  xA xB xC
1  0  1  0
2  1  0  0
3  1  0  0
4  0  1  0
5  0  0  1
6  0  1  0
7  0  1  0
8  0  0  1
9  0  1  0
```