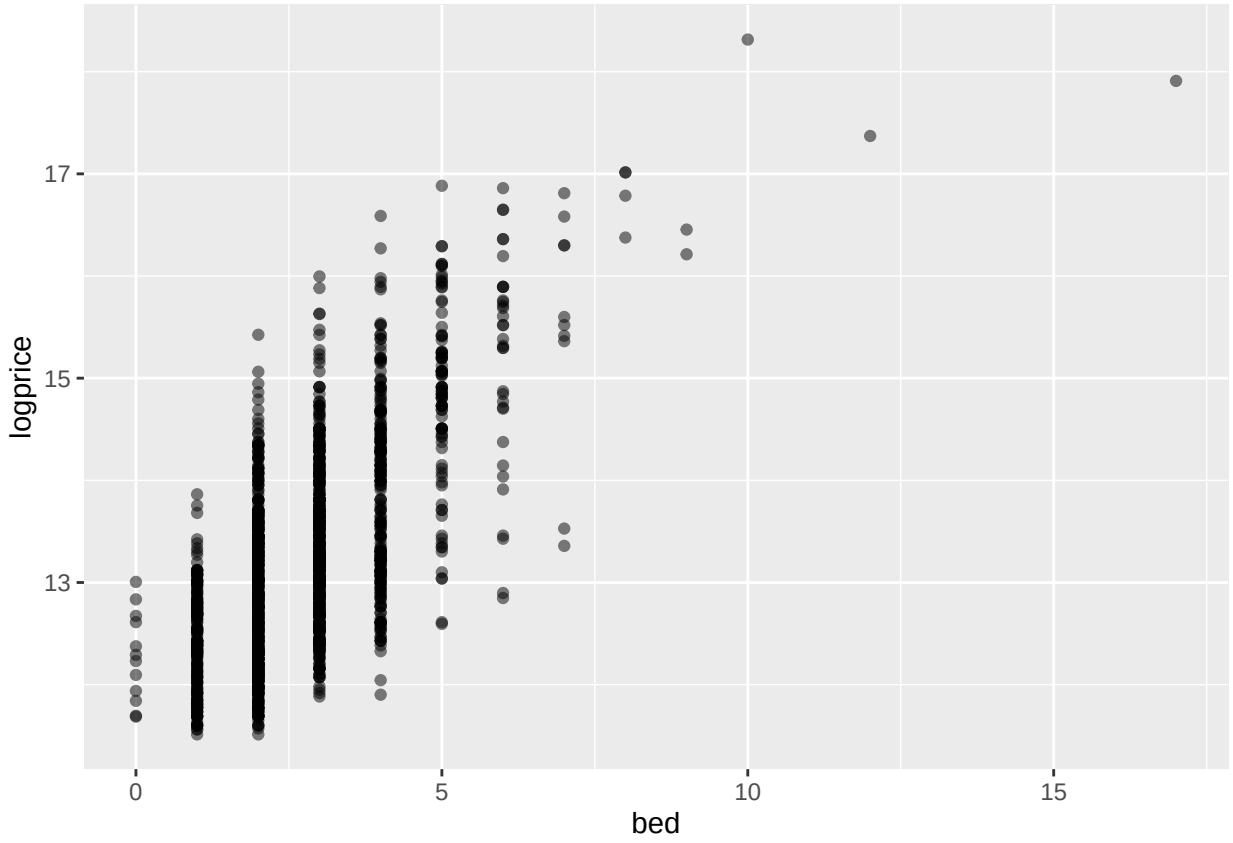
Activity 1

```
URL <- "http://andrewpbray.github.io/data/LA.csv"
LA <- read.csv(URL)
LA1 <- LA %>%
   mutate(logprice = log(price),
          logsqft = log(sqft))

m1 <- lm(logprice ~ logsqft + bed + city, data = LA1)
ggplot(data = LA1, mapping = aes(x = bed, y = logprice)) +
  geom_point(alpha = .5)
```



1. model seems to be a plane, except it would be 4 planes, 1 for each city.
2. The reference level for city is Beverly Hills.
3. The coefficient for bed, -.03010, seems to suggest that when you increase the number of beds while holding all the other variables constant, the price of your house goes down. This doesn't make too much sense to me because increasing the number of beds suggests increasing house size, which should increase house price. However, this is explained because if you hold square foot constant, if you increase number of beds, that means each bedroom will be smaller and more cramped, which would lower price. 2 houses of the same size in the same town. 5 bedrooms vs 4 bedrooms. It may be better to just pick one of these variables (bed/sqft).

Relationship between sqft and city ?

```
bedm <- lm(data = LA1, logprice ~ bed)
m2 <- lm(logprice ~ logsqft + bed +city + logsqft:city, data = LA1)
```

Santa Monica and Westwood may have different city/sqft relationships compared to Beverly Hills. Long Beach's interaction term is not significantly different than Beverly Hills.

Does the relationship between logsqft and logprice change depending on the number of beds?

```
m3 <- lm(logprice ~ logsqft + bed + logsqft:bed, data = LA1)
```

Non linear... Bendy plane; hard to interpret