code_laong

2025-10-08

## Categorical Inputs Practice

```r
library(ggplot2)

legos <- read.csv('https://vinnys-classes.github.io/data/legos_data.csv')
head(legos)
```

```
##   Item_Number                  Set_Name Theme Pieces Year Pages Minifigures
## 1       10859        My First Ladybird Duplo      6 2018     9          NA
## 2       10860        My First Race Car Duplo      6 2018     9          NA
## 3       10862      My First Celebration Duplo     41 2018     9          NA
## 4       10864 Large Playground Brick Box Duplo     71 2018    32           2
## 5       10867            Farmers' Market Duplo     26 2018     9           3
## 6       10870              Farm Animals Duplo     16 2018     8          NA
##     Packaging Unique_Pieces  Size amazon_price age
## 1         Box             5 Large        16.00   1
## 2         Box             6 Large         9.45   1
## 3         Box            18 Large        39.89   1
## 4 Plastic box            49 Large        56.69   2
## 5         Box            18 Large        36.99   2
## 6         Box            13 Large         9.99   2
```

```r
legos$Theme <- as.factor(legos$Theme)


ggplot(data = legos,
       aes(x = Theme,
           y = amazon_price)) +
  geom_point() +
  theme_classic()
```
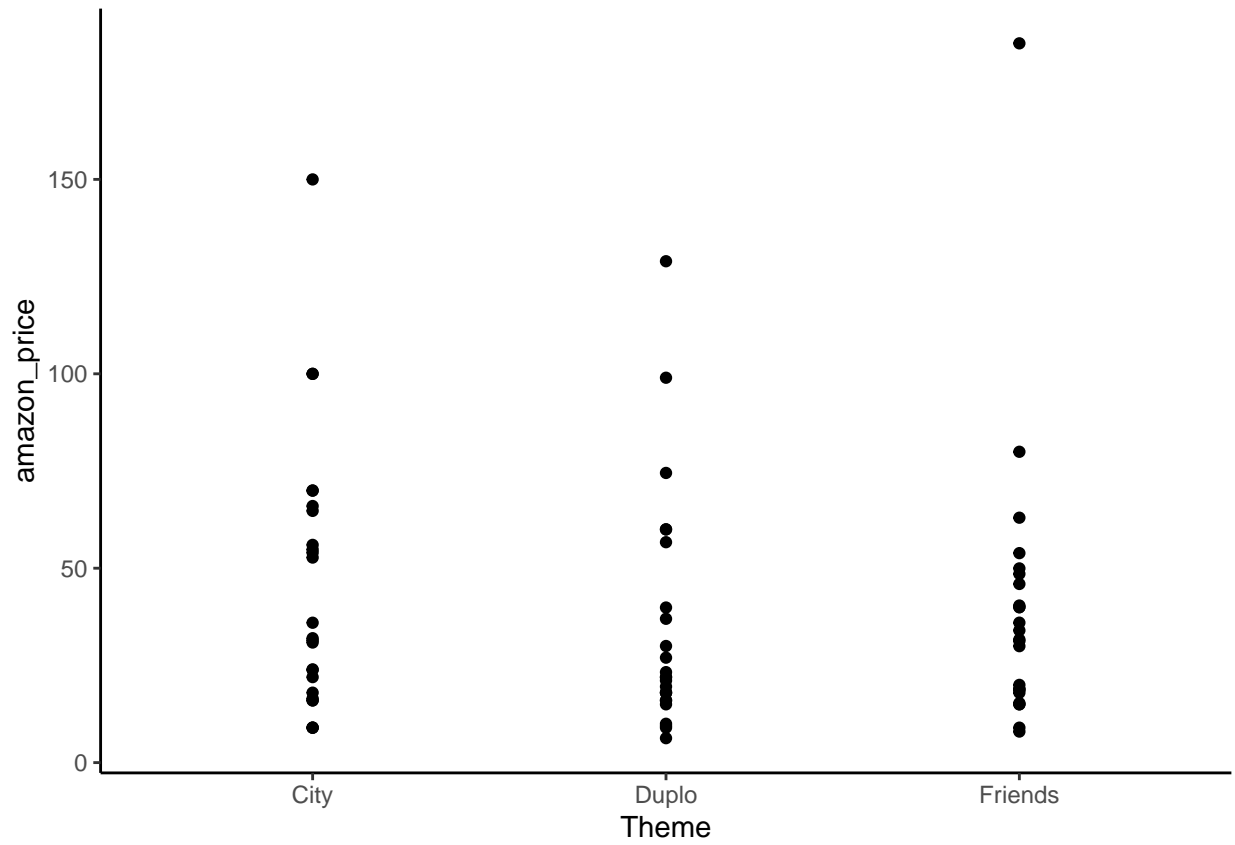
```
ggplot(data = legos,
       aes(x = Theme,
           y = amazon_price)) +
  geom_jitter(height = 0,
              width = .2) +
  theme_classic()
```
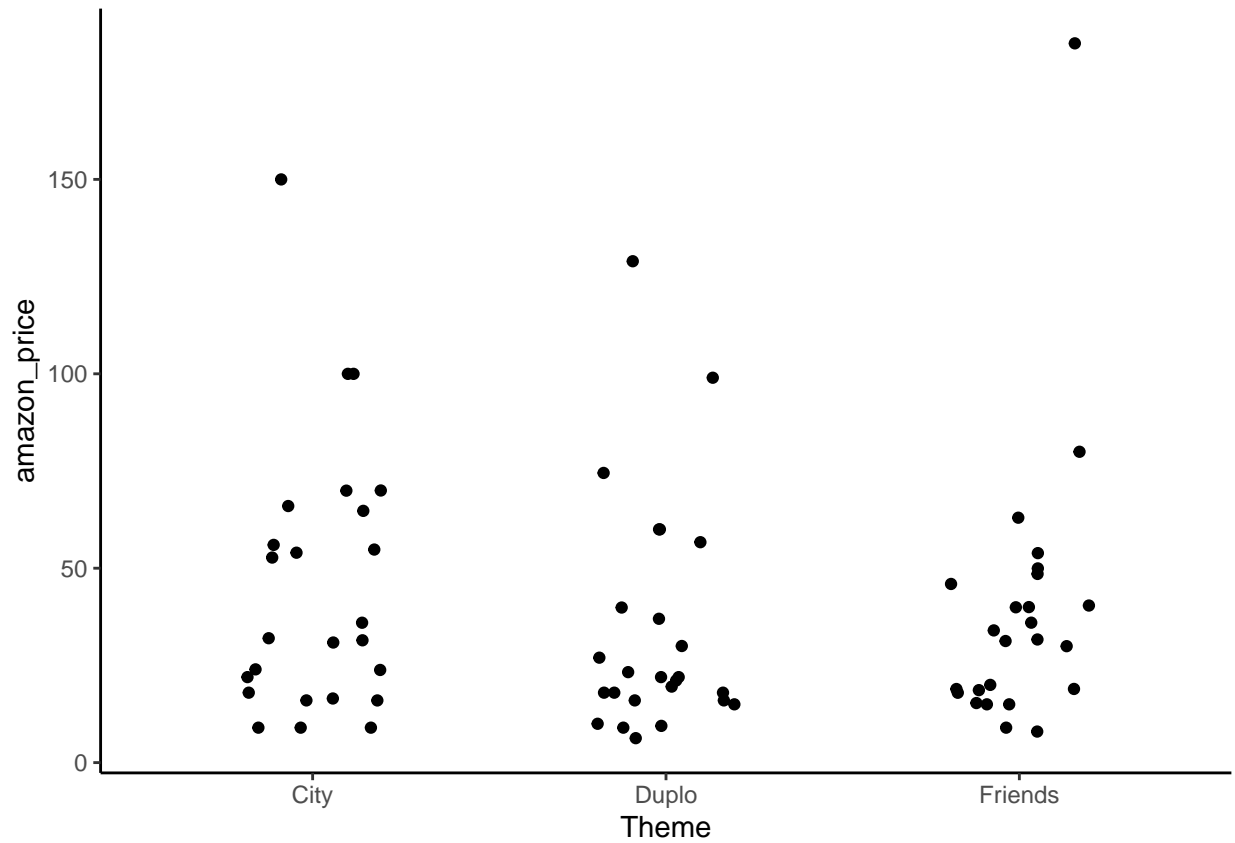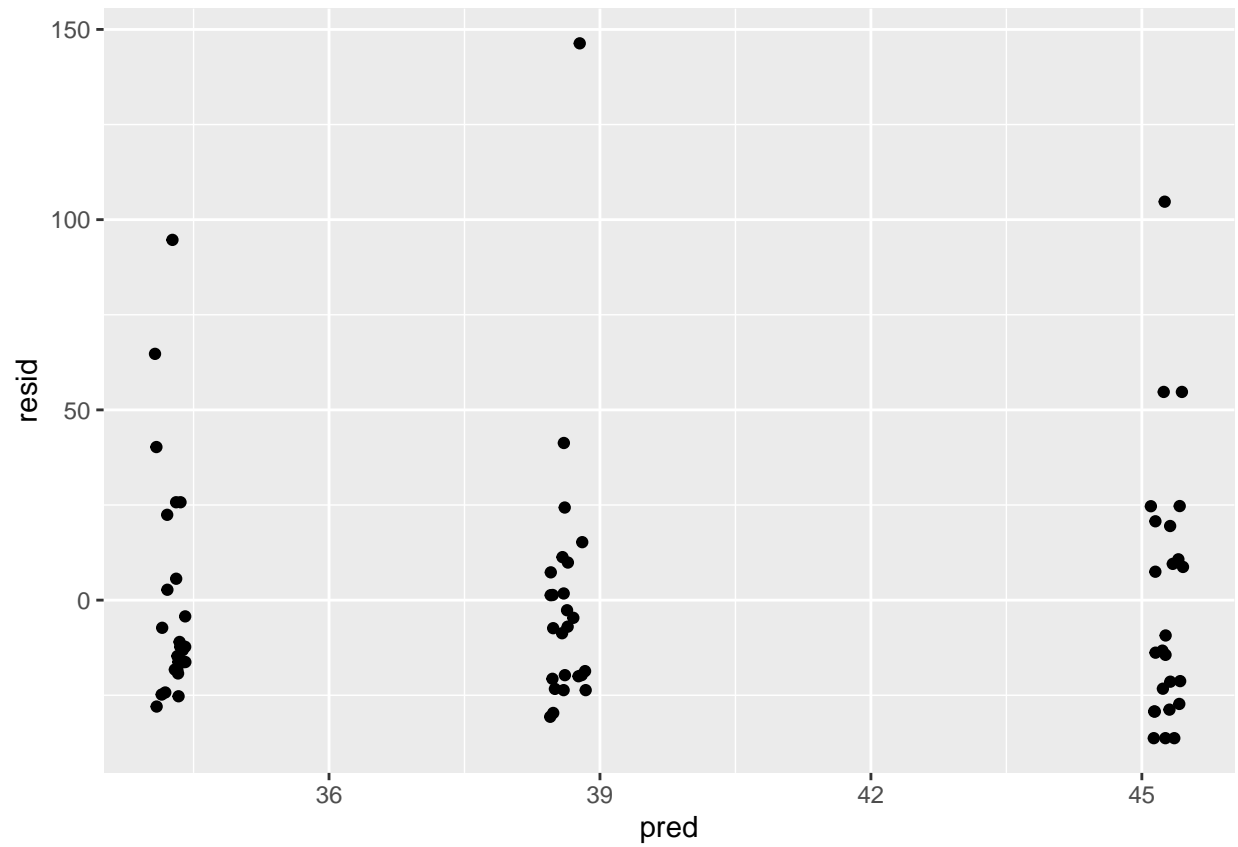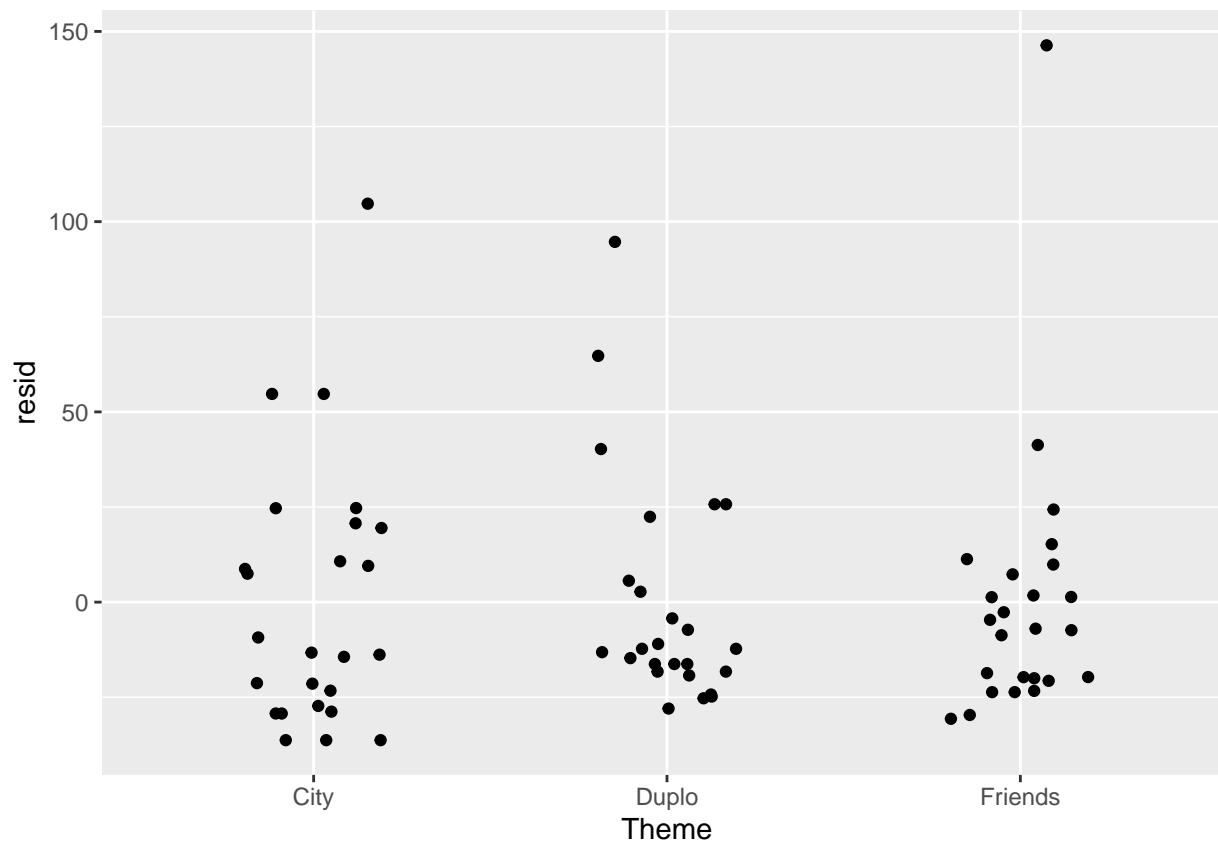
Fit the model

```r
mod <- lm(amazon_price ~ Theme,
          data = legos)


legos$resid <- resid(mod)
legos$pred <- predict(mod)


ggplot(data = legos,
       aes(x = pred,
           y = resid)) +
  geom_jitter(height = 0,
              width = .2)
```

```
ggplot(data = legos,
       aes(x = Theme,
           y = resid)) +
  geom_jitter(height = 0,
              width = .2)
```

Model Summary Stuff

```
summary(mod)
```

```
##
## Call:
## lm(formula = amazon_price ~ Theme, data = legos)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -36.28 -20.98 -10.99  10.31 146.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    45.270      6.689   6.768 2.95e-09 ***
## ThemeDuplo    -11.007      9.459  -1.164    0.248
## ThemeFriends   -6.620      9.459  -0.700    0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.44 on 72 degrees of freedom
## Multiple R-squared:  0.01871,    Adjusted R-squared:  -0.00855
## F-statistic: 0.6863 on 2 and 72 DF,  p-value: 0.5067
```

R^2: .01871

Find the outlier

```r
friends <- subset(legos, Theme == 'Friends')
head(friends)
```

```
##    Item_Number                 Set_Name   Theme Pieces Year Pages
## 26       41330 Stephanie's Soccer Practice Friends    119 2018    48
## 27       41333    Olivia's Mission Vehicle Friends    223 2018    84
## 28       41335           Mia's Tree House Friends    351 2018   120
## 29       41340           Friendship House Friends    722 2018   164
## 30       41353    Friends Advent Calendar Friends    500 2018     4
## 31       41356       Stephanie's Heart Box Friends     85 2019    32
##    Minifigures Packaging Unique_Pieces  Size amazon_price age      resid    pred
## 26           1       Box            78 Small        40.40   6    1.7508 38.6492
## 27           1       Box           106 Small        45.95   6    7.3008 38.6492
## 28           2       Box           151 Small        53.88   6   15.2308 38.6492
## 29           3       Box           309 Small       184.99   6  146.3408 38.6492
## 30          NA       Box           202 Small        34.00   6   -4.6492 38.6492
## 31           1      <NA>            36 Small        14.99   6  -23.6592 38.6492
```

```r
which.max(friends$amazon_price)
```

```
## [1] 4
```