

LEGO PRICES DATASET

By Angela - Robbie - Vicky

- 1,304 observations, 15 variables describing attributes of a Lego set posted on Brisket between Jan. 1, 2018 and Sept. 11, 2020.
- Investigate: How number of unique pieces, theme, and target age group are associated with the price of the lego set

OUR ASSUMPTIONS



V.S.



DATA INTERFERENCE

1. Missing data

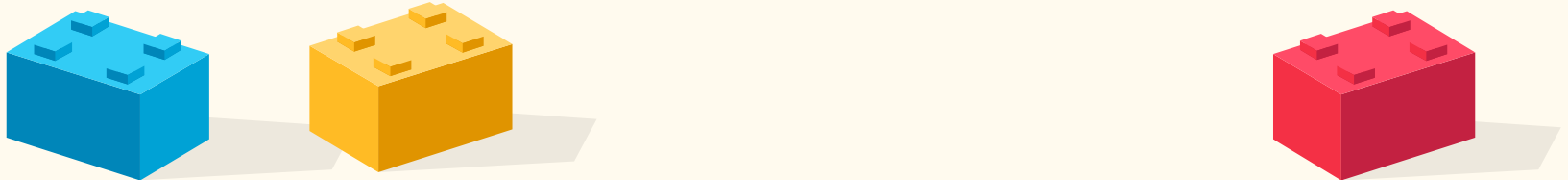
- unique_pieces: 46/1304 (3.5%)
- price: 239/1304 (18.33%)
- ages: 111/1304 (8.51%)
- theme: 270/1304 (20.71%)

2. Potential influence of missing data

- Mistakenly assume a normal distribution for an inference test
- Not representative of a bigger population

3. Data interference

- Collapse age groups
- Investigate only certain themes



ANOVA HYPOTHESIS TEST

AGE GROUPS - PRICE & THEMES - PRICE

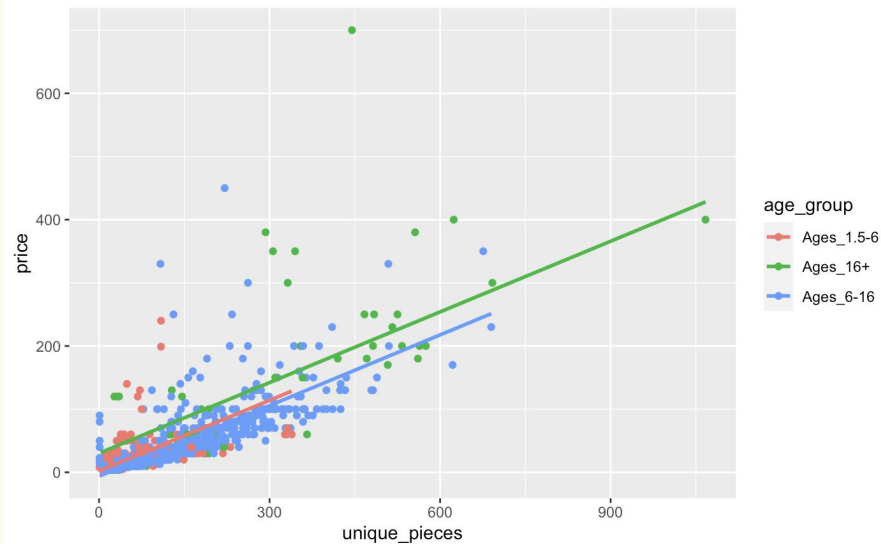
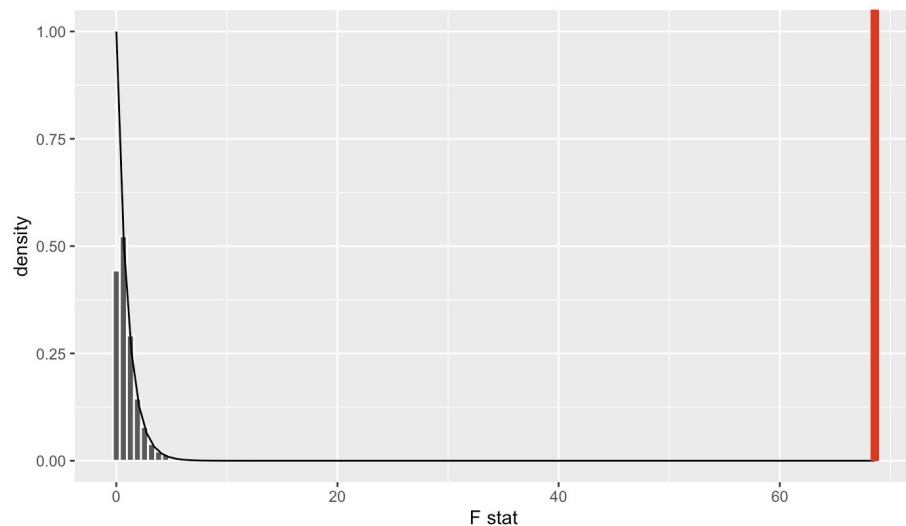
AGE GROUPS - PRICE

- Null Hypothesis: There is no difference among the mean prices of lego sets of different age groups
- Alternative hypothesis: There is a difference among the mean prices of lego sets of different age groups

THEMES - PRICE

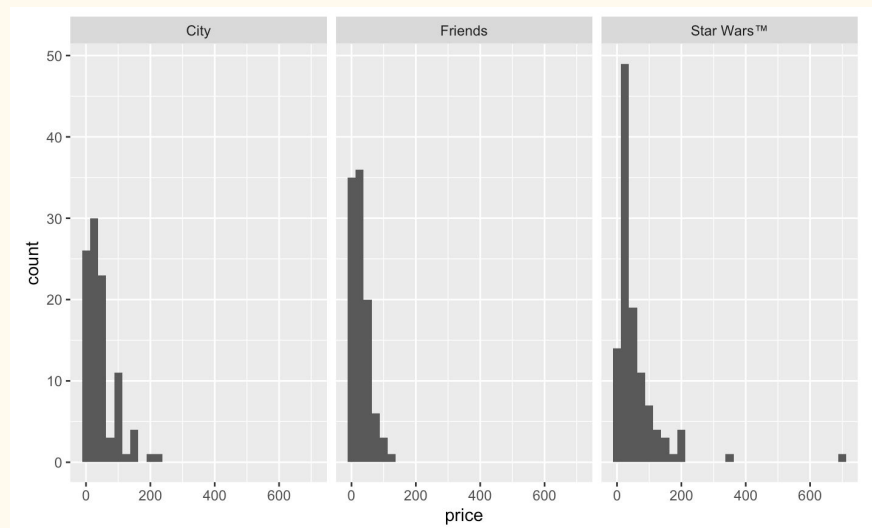
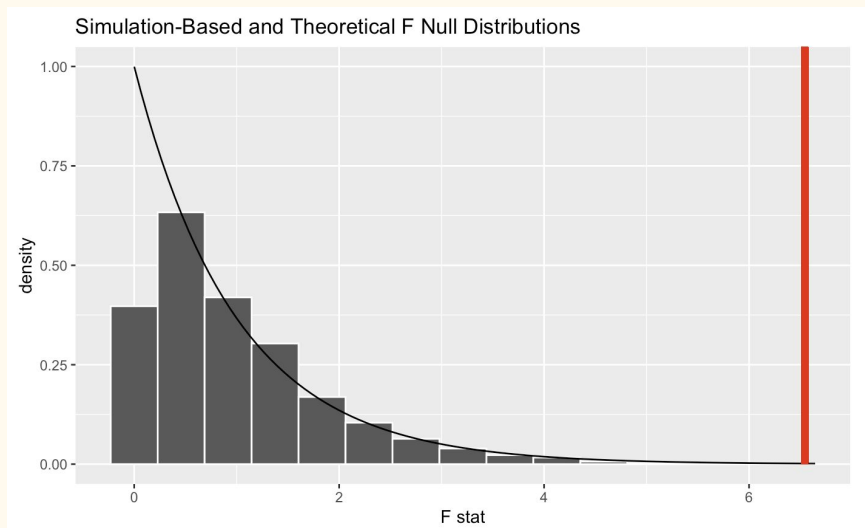
- Null Hypothesis: There is no difference among the mean prices of lego sets of different themes
- Alternative hypothesis: There is a difference among the mean prices of lego sets of different themes

Simulation-Based and Theoretical F Null Distributions



A tibble: 4 × 7

term <chr>	estimate <dbl>	std_error <dbl>	statistic <dbl>	p_value <dbl>	lower_ci <dbl>	upper_ci <dbl>
intercept	2.138	2.785	0.768	0.443	-3.327	7.604
unique_pieces	0.373	0.011	32.903	0.000	0.351	0.395
age_group: Ages_16+	28.046	5.413	5.181	0.000	17.424	38.668
age_group: Ages_6-16	-8.345	3.125	-2.670	0.008	-14.478	-2.212



term <chr>	estimate <dbl>	std_error <dbl>	statistic <dbl>	p_value <dbl>	lower_ci <dbl>	upper_ci <dbl>
intercept	-15.818	4.427	-3.573	0.000	-24.529	-7.108
unique_pieces	0.432	0.019	22.482	0.000	0.394	0.470
theme: Friends	-9.755	4.951	-1.970	0.050	-19.496	-0.013
theme: Star Wars™	13.047	4.803	2.716	0.007	3.596	22.498

4 rows