Lego

May 25, 2018

0.1 Introduction

Everyone loves Lego (unless you ever stepped on one). Did you know by the way that "Lego" was derived from the Danish phrase leg godt, which means "play well"? Unless you speak Danish, probably not.

In this project, we will analyze a fascinating dataset on every single lego block that has ever been built!

0.2 Reading Data

This comprehensive database of lego blocks is provided by Rebrickable. The data is available as csv files and the schema is shown below.

Let us start by reading in the colors data to get a sense of the diversity of lego sets!

In [2]:
```
# Import modules
import pandas as pd
import matplotlib.pyplot as plt
# Read colors data
colors = pd.read_csv('datasets/colors.csv')

# Print the first few rows
colors.head()
```

Out[2]:

	id	name	rgb	is_trans
0	-1	Unknown	0033B2	f
1	0	Black	05131D	f
2	1	Blue	0055BF	f
3	2	Green	237841	f
4	3	Dark Turquoise	008F9B	f

0.3 Exploring Colors

Now that we have read the colors data, we can start exploring it! Let us start by understanding the number of colors available.

In [3]:
```
# How many distinct colors are available?
# -- YOUR CODE FOR TASK 3 --
num_colors = colors.rgb.count()
num_colors
```

Out[3]: 135

0.4 Transparent Colors in Lego Sets

The colors data has a column named is_trans that indicates whether a color is transparent or not. It would be interesting to explore the distribution of transparent vs. non-transparent colors.

In [4]:
```
# colors_summary: Distribution of colors based on transparency
# -- YOUR CODE FOR TASK 4 -
colors_summary = colors.groupby('is_trans').count()
colors_summary
```

Out[4]:

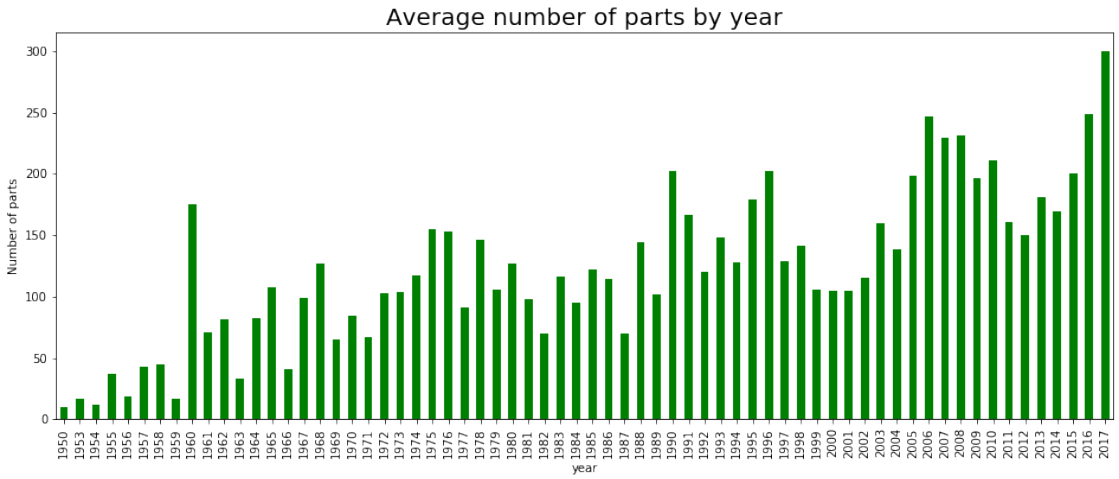
is_trans	id	name	rgb
f	107	107	107
t	28	28	28

0.5 Explore Lego Sets

Another interesting dataset available in this database is the sets data. It contains a comprehensive list of sets over the years and the number of parts that each of these sets contained.

Let us use this data to explore how the average number of parts in lego sets has varied over the years.

In [5]:
```
%matplotlib inline
# Read sets data as `sets`
sets = pd.read_csv('datasets/sets.csv')
# Create a summary of average number of parts by year: `parts_by_year`
parts_by_year = sets.groupby('year')['num_parts'].mean()
# Plot trends in average number of parts by year
fig, ax = plt.subplots(figsize=(16,6))
parts_by_year.plot.bar(ax=ax,color = 'green')
ax.set_title("Average number of parts by year", fontsize= 20)
ax.set_ylabel("Number of parts")
plt.show()
```



0.6 Lego Themes Over Years

Lego blocks ship under multiple themes. Let us try to get a sense of how the number of themes shipped has varied over the years.

```
In [6]: # themes_by_year: Number of themes shipped by year
        # -- YOUR CODE HERE --
        themes_by_year = sets[['year', 'theme_id']].groupby('year', as_index = False).agg({"the
        themes_by_year.head()
```

Out[6]:

	year	theme_id
0	1950	2
1	1953	1
2	1954	2
3	1955	4
4	1956	3

0.7 Wrapping It All Up!

Lego blocks offer an unlimited amoung of fun across ages. We explored some interesting trends around colors, parts and themes.