CZ4124/SC4024 Assignment #1 - Summary Report

Name: Khant Nyi Nyi **Title**: Short History of Lego sets

1. Objectives

The primary objective of the project is to explore how different aspects of Lego sets may have changed over time. The aspects being explored in the videos were as shown below:

- The cumulative number of sets produced per theme across the year: the Lego group has produced sets from over 130 themes throughout history and the total production of sets per theme is explored through a racing bar chart.
- Lego themes can be sub-divided into two categories: Original and licensed. The proportions of these sub-categories are explored using a tree map and doughnut map.
- The size of Lego sets over the year, measured in piece count, is explored using basic bar charts.

2. Novelty

One original visualisation used in the video was the racing bar charts. This video sequencing technique was thought to offer a unique way of visualising quantitative data over time. As this technique was not covered in the course, various tutorials had to be looked up to find a proper method of creating such videos. With new visualisation techniques, also comes a need for more data wrangling as cumulative sum of data points were needed to be aggregated for the creation of the racing bar charts video.

The sales data visualisation can be seen on the original Statista website. However, in this video presentation, data point labelling thought to be redundant were removed while more appropriate data labelling was added, and the colour scheme was modified to fit the Lego colour palette better. An accompanying choropleth map was also created and added for easier visualisation of the regions division.

3. Technical Challenges and Innovation

Data exploration is mostly done through Jupyter notebook with seaborn and matplotlib as plotting tools for exploration of parameters such as the size of window in rolling mean. The majority of the visualisations seen in the video are done using Flourish, a web application where various plots can be created and modified using a GUI instead of coding out the parameters. Data is exported after cleaning and wrangling in the Jupyter notebook and imported into Flourish for plot parameters adjustment and beautification for the final visuals seen in the video.

The main challenges of this video project come mostly from data wrangling and less from creating the visualisations themselves. The initial dataset contained 11 data schemas as seen in *Figure 1*. After preliminary exploration and formulation of guiding questions, the number of data schemas required for the project was narrowed down to 5. Even then, the merging of data schemas with the given relationships in *Figure 1* to create appropriate data frames to address the guiding questions proved to be a challenge as well.

Another major challenge faced was finding an appropriate visualisation for the sales growth by region. As geospatial visualisations tend to be static, there was difficulty in finding an appropriate form of time series of geospatial data. Eventually, a stacked bar chart was used to visualise the time element of the data while an equirectangular choropleth map was used for referencing of the regions involved.

4. References

Data sources:

- Lego sets data: Free LEGO Catalog Database Downloads | Rebrickable - Build with LEGO
  - The Rebrickable catalogue database provides a comprehensive structured datasets of the properties of all the Lego sets produced.
  - Downloaded in September 2023 but only data from 2022 and before are explored in the project.
- Lego sales by region: LEGO: net sales by region worldwide 2022 | Statista
  - A succinct excel file containing the sales data is downloaded from this link.
  - The original source of the data is originally from The Lego Annual Report 2022

Other links:

- Flourish web app: Flourish | Data Visualization & Storytelling
  - This web application is used to create most of the visuals seen in the video
- Custom Choropleth Map maker: Simple Choropleth Map Maker (Previously: Alternate History Editor) (yulin-w.github.io)
  - This web app is used to create an accompanying region mapping for the sales data.
- Tutorial on how to create a racing bar chart: How To Make Racing Bar Graph Comparison Videos (Full Tutorial) - YouTube
- Brickset Website: Home | Brickset
  - The Brickset website provides comprehensive information of all Lego sets produced.
  - The Brickset website is used extensively during the exploration of the data set of verify the nature and properties of some data points (specific Lego sets)

Figures and Tables

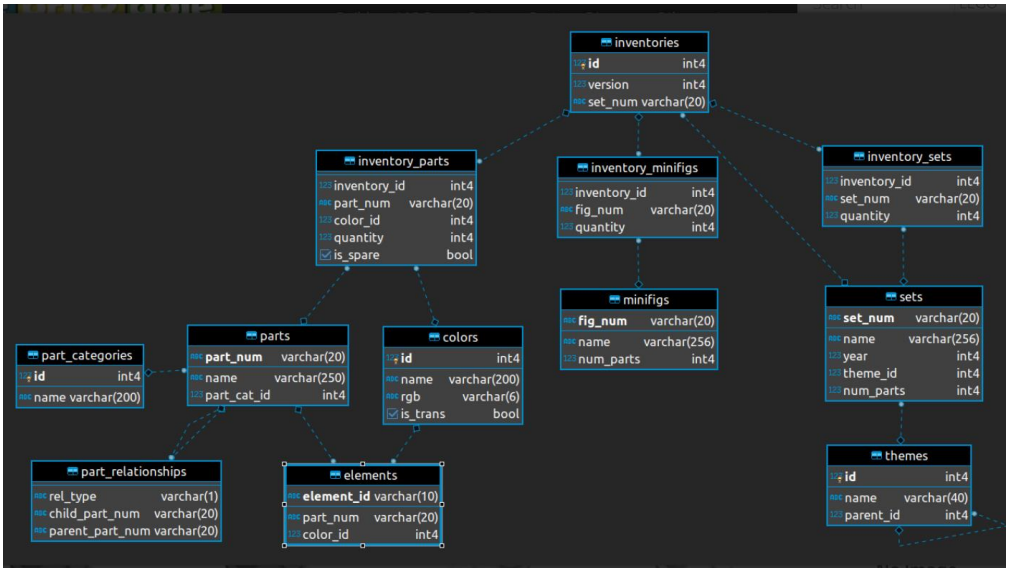

Figure 1: Datasets schemas and relationships