

Group 30

DATA WRANGLING PROJECT

By:
Karolina Hajkova,
Matt Hrkal,
Antal Deurloo,
Malgorzata Zdych

Introduction

The Research questions

What is the correlation between a country's various indexes and the number of medals awarded during the Olympic Games?

How do different statistics of countries correlate to how successful they perform at the Olympic Games?



Data sources

**Tokyo 2020 Olympics
Medals (Kaggle)**

**Population by country 2020
(Kaggle)**

**Global Corruption
Perceptions Index for 10
years (Kaggle)**

**Alcohol consumption
around the world (Kaggle)**

**Population by Country
(Kaggle)**



Data Wrangling



Data Overview

Data cleaning

Data preparation

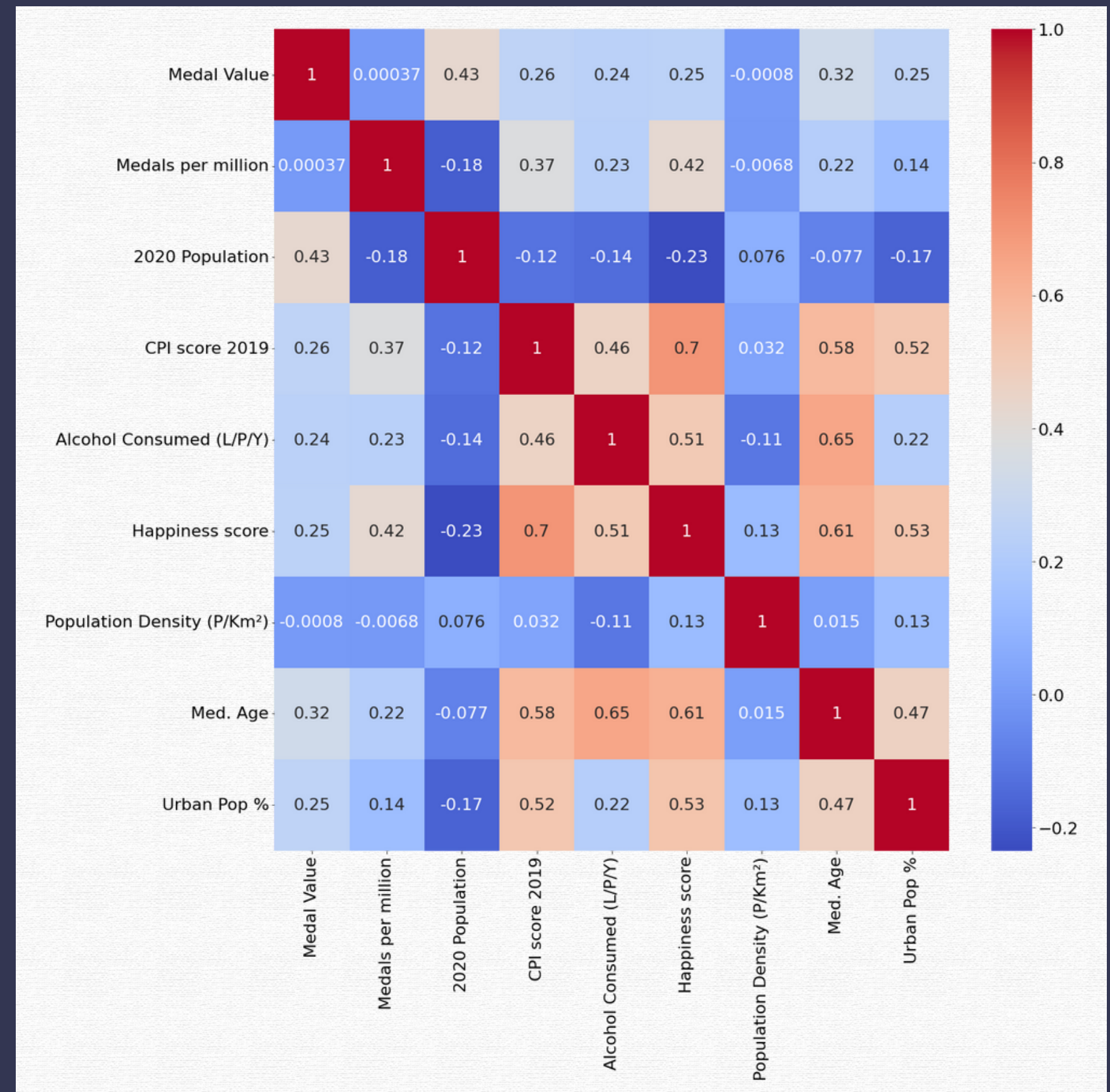
Data analysis

Data Wrangling

- 1 Standardizing country names
- 2 Translating country names to country codes
- 3 Handling non-standard data
- 4 Developing the medal scoring system
- 5 Merging data frames

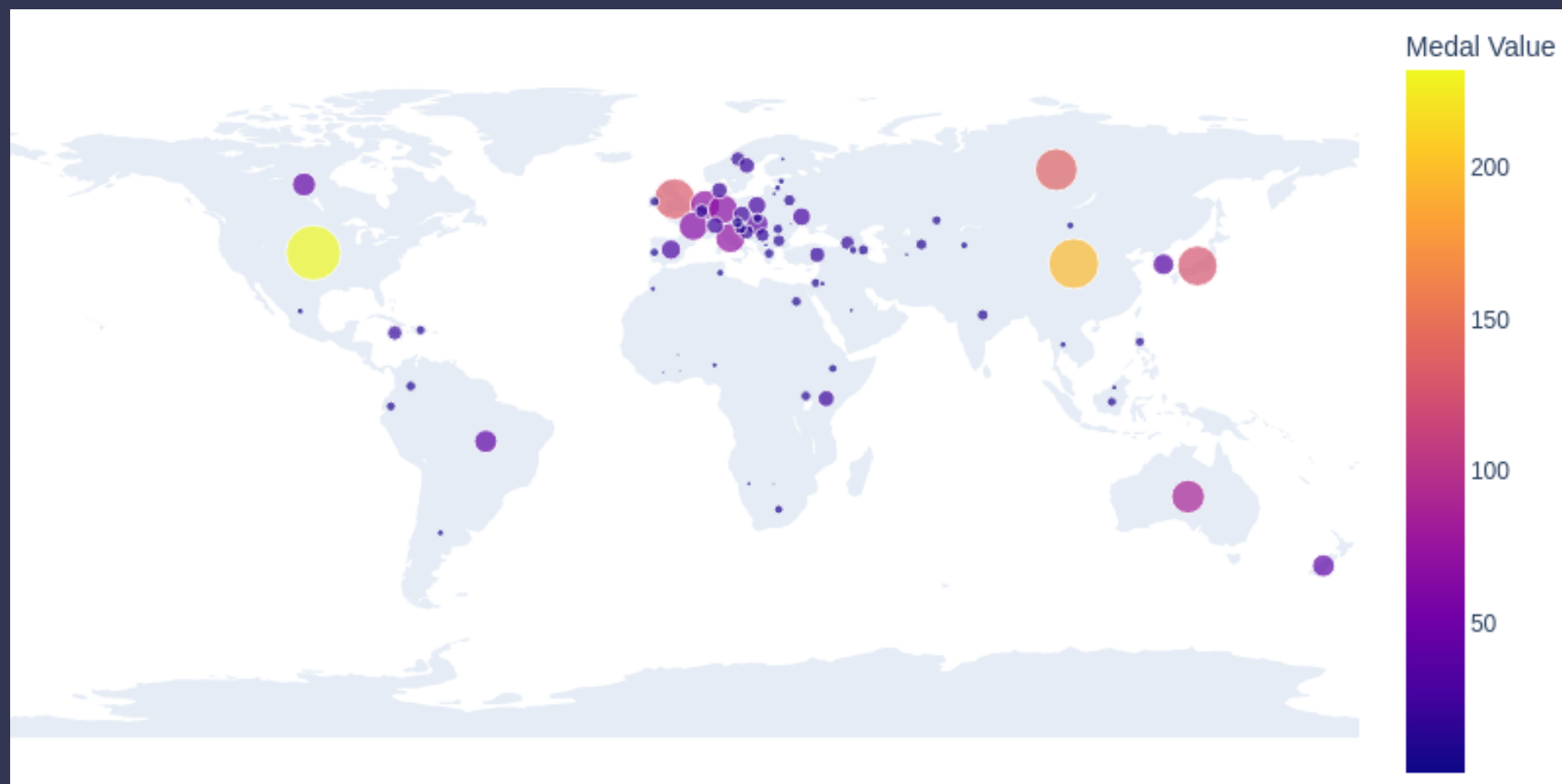
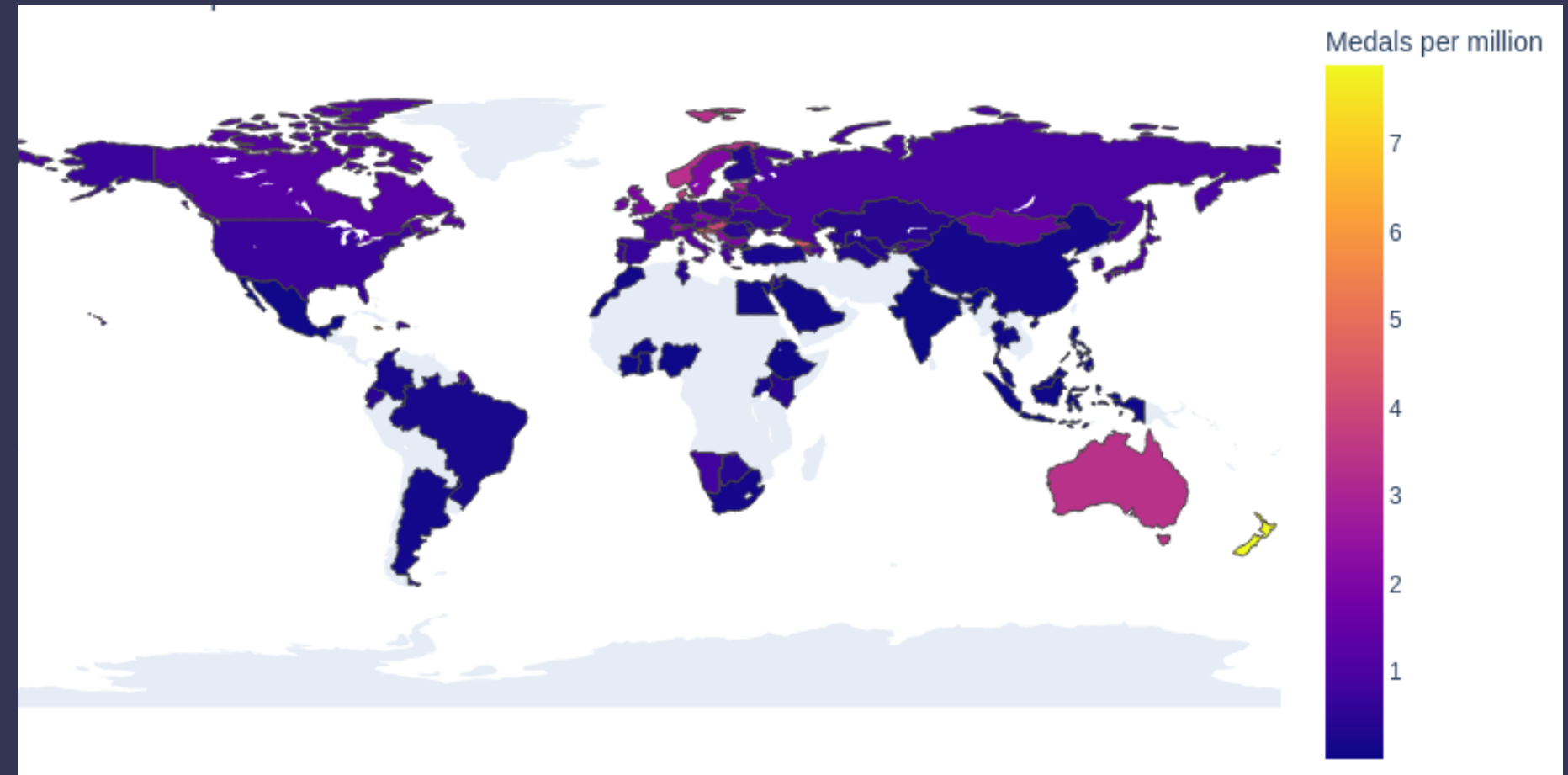
Data analysis

- **Positive correlations:**
Medals/Million & CPI
Population & Medal
Medals/Million & Happiness



Data analysis

- **Bubble & Choropleth maps:** Comparing medals & population.



- **Small countries:** Potential to succeed.

Conclusion

Correlation doesn't always imply causality; further investigation is needed.

Other factors such as income, education, and culture could be at play.

Analysis highlights possible relationships between variables and the need to consider multiple factors when exploring correlations.





THANK YOU