Exploring Unicorn Companies

March 2, 2022

The dataset used for this challenge is titled 'unicorn.csv'.

What topic does the dataset cover?

This dataset was pulled from https://www.kaggle.com/datasets/deepcontractor/unicorn-companies-dataset. A unicorn company, or unicorn startup, is a private company with a valuation over $1 billion. As of March 2022, there are 1,000 unicorns around the world. Popular former unicorns include Airbnb, Facebook and Google. Variants include a decacorn, valued at over $10 billion, and a hectocorn, valued at over $100 billion.

Variables

- Company: Company name
- Valuation: The company valuation in billions of dollars.
- Country: Country of company.
- City: City of company.
- Industry: Industry of the company
- Select_investors: Who invested in the company
- Founded_year: Year the company was founded
- Financial_stage: Financial stage of the company.
- Investors_count: Number of investors
- Deal_terms: Number of terms in investment deals

Assignment

1. Visualize and describe the distribution for unicorn company valuations.
2. Which five countries have the largest number of unicorns?
3. Looking at the five countries in the previous question, do they differ in valuations on average? Use pariwise boxplots to support your answer.
4. Which five industries have the most amount of unicorns. Which industry would you most like to work in?
5. Looking at the five industries found in the last problem, do they differ in valuations on average? Use pairwise boxplots to support your answer.

6. Which 5 companies have the highest valuations? Which company would you most like to work for?

7. Has the number of unicorn companies changed over time?

8. If you just sold a company for $1 billion, would you spend the money on?

Solution:

```
### Load packages and dataset
library(tidyverse)
library(ggplot2)
uni <- read.csv('/data/datasets/unicorn.csv')

### Take a look at the dataset
glimpse(uni)

### Question 1: Describe the distribution for company valuations
uni %>% ggplot(aes(x=Valuation)) +
  geom_histogram(color='black', fill='green')+
  theme_classic()+
  labs(title='Distribution for company valulations in billions of dollars',
       x="Valuation in Billions of dollars")

### Question 2: Which 5 countries have the largest number of Unicorn companies?
uni %>% group_by(Country) %>% summarize(count=n()) %>% arrange(desc(count))

### Question 3: DO valuations differ by the top 5 countries?
uni %>% filter(Country=='United States' |
                 Country == "China" |
                 Country == "India" |
                 Country == "United Kingdom" |
                 Country == "France") %>%
  ggplot(aes(y=Valuation, fill=Country)) +
  geom_boxplot() +
  theme_classic()

### Question 4: Which 4 industries have the most amount of unicorns?
uni %>% group_by(Industry) %>% summarize(count=n()) %>% arrange(desc(count))

### Question 5: DO valuations differ by the top 5 Industries?
uni %>%
  filter(Industry=='Fintech' |
           Industry == "Internet software & services" |
           Industry == "E-commerce & direct-to-consumer" |
           Industry == "Artificial intelligence" |
           Industry == "Health") %>%
  ggplot(aes(y=Valuation, fill=Industry)) +
  geom_boxplot() +
  theme_classic()

### Question 6: Which 5 companies have the highest valuations?
uni %>% arrange(desc(Valuation)) %>%
  select(Company) %>%
  head(10)

### Question 7: Has the number of unicorn companies created changed over time?
uni %>% ggplot(aes(x=Founded_year)) +
  geom_bar() + theme_classic()
```