Data Science - Practice 4.5

Figure 1: pokemon

이 데이터는 포켓몬 7세대까지 등장한 모든 801종의 포켓몬에 대한 정보이다. 이 데이터에는 기본 능력치, 다른 타입에 대한 상성, 키, 몸무게, 종류, 알에서 부화하기 위한 걸음수, 경험치, 능력치 등을 포함한다.

This dataset contains information on all 801 Pokemon from all Seven Generations of Pokemon. The information contained in this dataset include Base Stats, Performance against Other Types, Height, Weight, Classification, Egg Steps, Experience Points, Abilities, etc.

- name: The English name of the Pokemon
- pokedex_number: The entry number of the Pokemon in the National Pokedex
- percentage_male: The percentage of the species that are male. Blank if the Pokemon is genderless.
- type1: The Primary Type of the Pokemon
- type2: The Secondary Type of the Pokemon
- classification: The Classification of the Pokemon as described by the Sun and Moon Pokedex
- height_m: Height of the Pokemon in metres
- weight_kg: The Weight of the Pokemon in kilograms
- capture_rate: Capture Rate of the Pokemon
- baseeggsteps: The number of steps required to hatch an egg of the Pokemon
- abilities: A stringified list of abilities that the Pokemon is capable of having
- experience_growth: The Experience Growth of the Pokemon
- base_happiness: Base Happiness of the Pokemon
- against_?: Eighteen features that denote the amount of damage taken against an attack of a particular type
- hp: The Base HP of the Pokemon
- attack: The Base Attack of the Pokemon
- defense: The Base Defense of the Pokemon
- sp_attack: The Base Special Attack of the Pokemon
- sp_defense: The Base Special Defense of the Pokemon
- speed: The Base Speed of the Pokemon
- generation: The numbered generation which the Pokemon was first introduced
- is_legendary: Denotes if the Pokemon is legendary.

모든 문제에 대하여 R code와 결과, 결과에 대한 해석을 포함하세요.

For all questions include explanation of the process and the result in your report as well as R code.

< Question 1 >

R 작업 환경에 데이터를 불러온 후 데이터에 대한 기본적인 정보를 파악하고 기술하시오.

Load the dataset into R for data exploration. Describe the basic overview of the dataset.

Sample Result

```
## Rows: 801
## Columns: 40
## $ abilities         <chr> "['Overgrow', 'Chlorophyll']", "['Overgrow', 'Chloro~
## $ against_bug       <dbl> 1.00, 1.00, 1.00, 0.50, 0.50, 0.25, 1.00, 1.00, 1.00~
## $ against_dark      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ against_dragon    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ against_electric  <dbl> 0.5, 0.5, 0.5, 1.0, 1.0, 2.0, 2.0, 2.0, 2.0, 1.0, 1.~
## $ against_fairy     <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1.0, 1.0, 1.~
## $ against_fight     <dbl> 0.50, 0.50, 0.50, 1.00, 1.00, 0.50, 1.00, 1.00, 1.00~
## $ against_fire      <dbl> 2.0, 2.0, 2.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2.0, 2.~
## $ against_flying    <dbl> 2.0, 2.0, 2.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 2.~
## $ against_ghost     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0~
## $ against_grass     <dbl> 0.25, 0.25, 0.25, 0.50, 0.50, 0.25, 2.00, 2.00, 2.00~
## $ against_ground    <dbl> 1.0, 1.0, 1.0, 2.0, 2.0, 0.0, 1.0, 1.0, 1.0, 0.5, 0.~
## $ against_ice       <dbl> 2.0, 2.0, 2.0, 0.5, 0.5, 1.0, 0.5, 0.5, 0.5, 1.0, 1.~
## $ against_normal    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ against_poison    <dbl> 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.~
## $ against_psychic   <dbl> 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1~
## $ against_rock      <dbl> 1, 1, 1, 2, 2, 4, 1, 1, 1, 2, 2, 4, 2, 2, 2, 2, 2, 2~
## $ against_steel     <dbl> 1.0, 1.0, 1.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.~
## $ against_water     <dbl> 0.5, 0.5, 0.5, 2.0, 2.0, 2.0, 0.5, 0.5, 0.5, 1.0, 1.~
## $ attack            <int> 49, 62, 100, 52, 64, 104, 48, 63, 103, 30, 20, 45, 3~
## $ base_egg_steps    <int> 5120, 5120, 5120, 5120, 5120, 5120, 5120, 5120, 5120~
## $ base_happiness    <int> 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, ~
## $ base_total        <int> 318, 405, 625, 309, 405, 634, 314, 405, 630, 195, 20~
## $ capture_rate      <chr> "45", "45", "45", "45", "45", "45", "45", "45", "45"~
## $ classfication     <chr> "Seed Pokemon", "Seed Pokemon", "Seed Pokemon", "Liz~
## $ defense           <int> 49, 63, 123, 43, 58, 78, 65, 80, 120, 35, 55, 50, 30~
## $ experience_growth <int> 1059860, 1059860, 1059860, 1059860, 1059860, 1059860~
## $ height_m          <dbl> 0.7, 1.0, 2.0, 0.6, 1.1, 1.7, 0.5, 1.0, 1.6, 0.3, 0.~
## $ hp                <int> 45, 60, 80, 39, 58, 78, 44, 59, 79, 45, 50, 60, 40, ~
## $ name              <chr> "Bulbasaur", "Ivysaur", "Venusaur", "Charmander", "C~
## $ percentage_male   <dbl> 88.1, 88.1, 88.1, 88.1, 88.1, 88.1, 88.1, 88.1, 88.1~
## $ pokedex_number    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1~
## $ sp_attack         <int> 65, 80, 122, 60, 80, 159, 50, 65, 135, 20, 25, 90, 2~
## $ sp_defense        <int> 65, 80, 120, 50, 65, 115, 64, 80, 115, 20, 25, 80, 2~
## $ speed             <int> 45, 60, 80, 65, 80, 100, 43, 58, 78, 45, 30, 70, 50,~
## $ type1             <chr> "grass", "grass", "grass", "fire", "fire", "fire", "~
## $ type2             <chr> "poison", "poison", "poison", "", "", "flying", "", ~
## $ weight_kg         <dbl> 6.9, 13.0, 100.0, 8.5, 19.0, 90.5, 9.0, 22.5, 85.5, ~
## $ generation        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ is_legendary      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

< Question 2 >

포켓몬은 타입에 따라서 다른 성질을 가진다.

포켓몬의 타입(type1) 에 따른 능력치 (attack, defense, speed, hp, sp_attack, sp_defense)의 특징을 기술하고,

가능하다면 적합한 plot을 사용하여 시각화 하여 보아라.

Pokemons are known to have different characteristics for their type.

Charaterize the types of Pokemon with their Primary classification(type1) in terms of their abilities including attack, defense, speed, hp, sp_attack, sp_defense.

Use a proper plot to demonstrate characteristics of each group if necessary.

< Question 3 >

포켓몬은 다양한 키와 몸무게 수치를 가지고 있다.

덩치카 큰 포켓몬이 방어력이 높다고 할 수 있는가?

덩치가 작은 포켓몬이 속도가 빠르다고 할 수 있는가?

덩치가 큰 포켓몬과 작은 포켓몬을 그룹을 나누어 비교해보자. (능력적으로 차이점이 있는지)

가능하다면 이 문제를 설명하기 위하여 적합한 plot으로 시각화 해보아라.

Pokemons are widely different in their size (height and weight).

Can we say bigger pokemons are stronger in their defense?

Can we say smaller pokemons are faster in their speed?

Compare groups of small and large pokemons in various aspects of their abilities.

Use a proper plot to demonstrate characteristics of each group if necessary.

< Question 4 >

각 세대별로 다양한 포켓몬이 소개되었다.

세대별 포켓몬의 특징을 주어진 데이터를 통하여 설명해보아라.

각 세대별로 어떤 특징을 찾을 수 있을까?

Many unique pokemons were showcased in 7 different generations.

Can you characterize each generation using the dataset.

What are the unique features of each generation?

< Question 5 >

당신이 피카츄와 대결해야한다고 가정해보자.

이때 능력치와 성질을 기반으로 피카츄와 대결하기에 유리한 10개의 포켓몬을 찾아보자.

왜 당신이 고른 포켓몬이 피카츄에 유리한지, 데이터셋을 통해 찾아낸 정보를 기반으로 설명하라.

Assume that you are fighting against Pikachu which is main pokemon of this series.

Find top 10 pokemons which are most advantageous to Pikachu in their ability and characteristics.

Explain why they are advantageous with findings from the dataset.

< Question 6 >

포켓몬 중에는 매우 희귀한 전설의 포켓몬 (legendary) 이 있다.
전설의 포켓몬과 일반 포켓몬의 가장 큰 차이점은 무엇인가.
어떤 특징을 통해서 포켓몬이 전설의 포켓몬인지 구분할 수 있을지 설명해보자.
There are very rare legendary pokemons.
What are the main difference between legendary pokemons and normal pokemons?
Explain the way we figure out if certain pokemon is legendary. how can we distinguish?

< bonus question >

포켓몬의 한국 이름과 영문명은 다르다. 한국 포켓몬 팬은 영어 이름보다 한국 이름에 친숙하다.
'Korean_name'이라는 포켓몬의 한국 번역 이름을 담은 새로운 변수를 만들어라.
한국어 이름을 추가하기 위해서 쉽고 간편한 방법이 있다면 설명해보자.
Korean fans of Pokemon are more familiar with Korean name of pokemon than english name.
Add new column 'Korean_name' contains Korean translated names of Pokemon.
Explain if there is efficient and easy way to do the job.