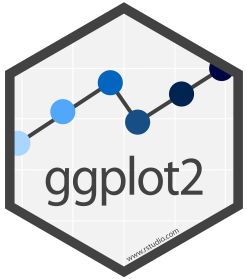
# INTRO TO THE **TIDYVERSE**

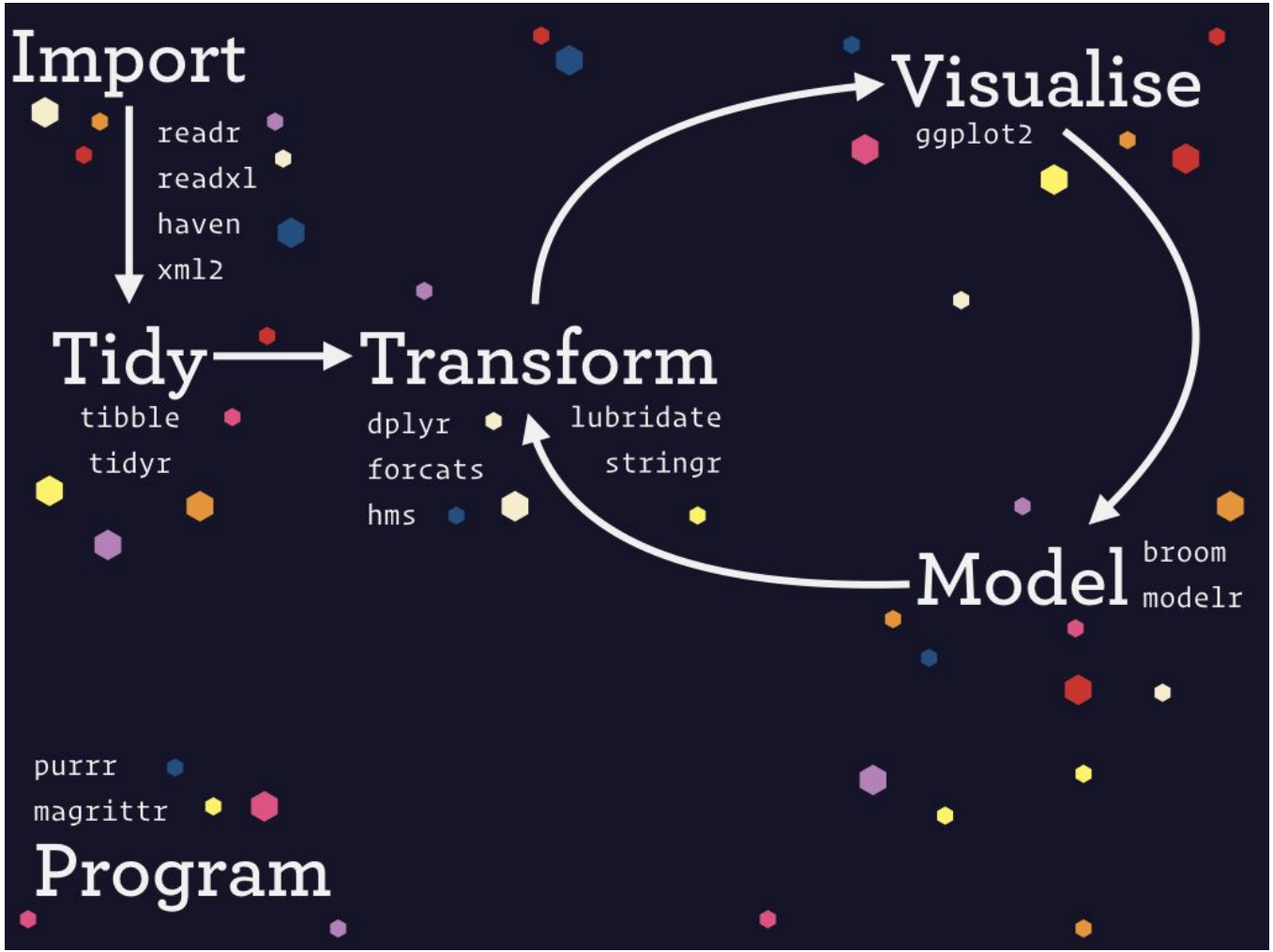
DATA VISUALIZATION USING




OMAYMA SAID

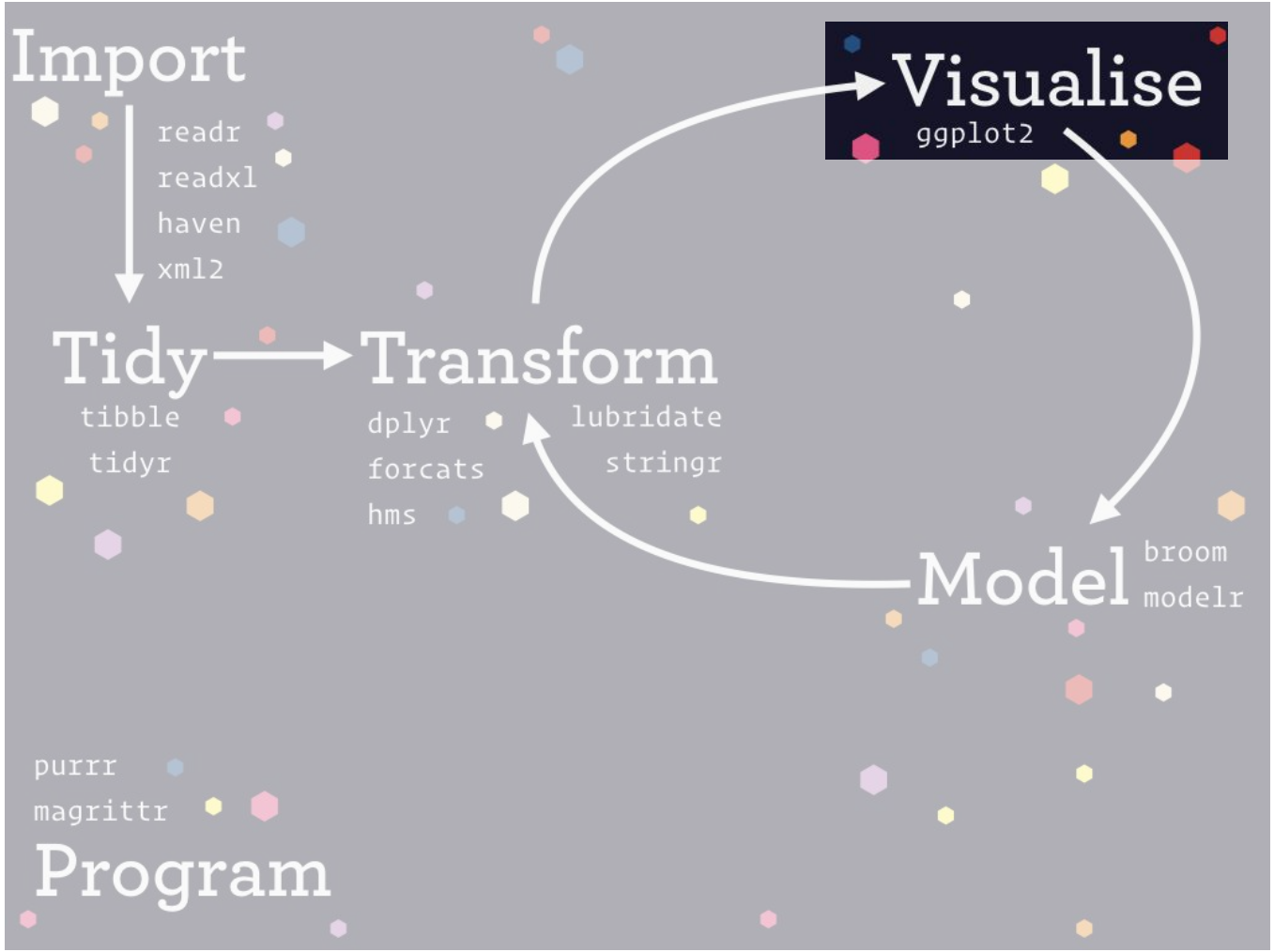
OmaymaS

# The Tidyverse

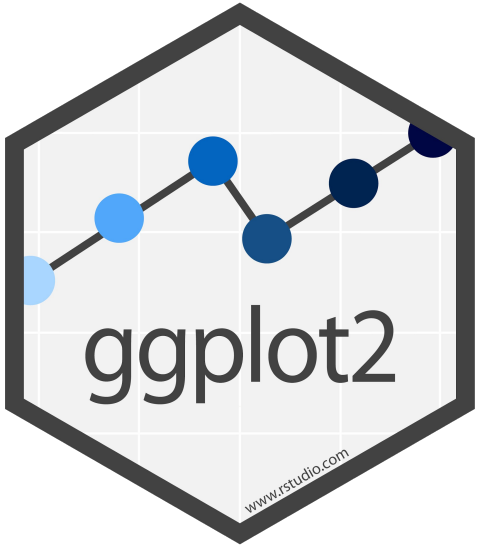



Source: https://imgur.com/a/l7fNwP1

# The Tidyverse




Source: https://imgur.com/a/l7fNwP1




Based on the grammar of graphics

# DATASETS

## PART 1

### gapminder

Main Source: gapminder dataset

https://github.com/jennybc/gapminder

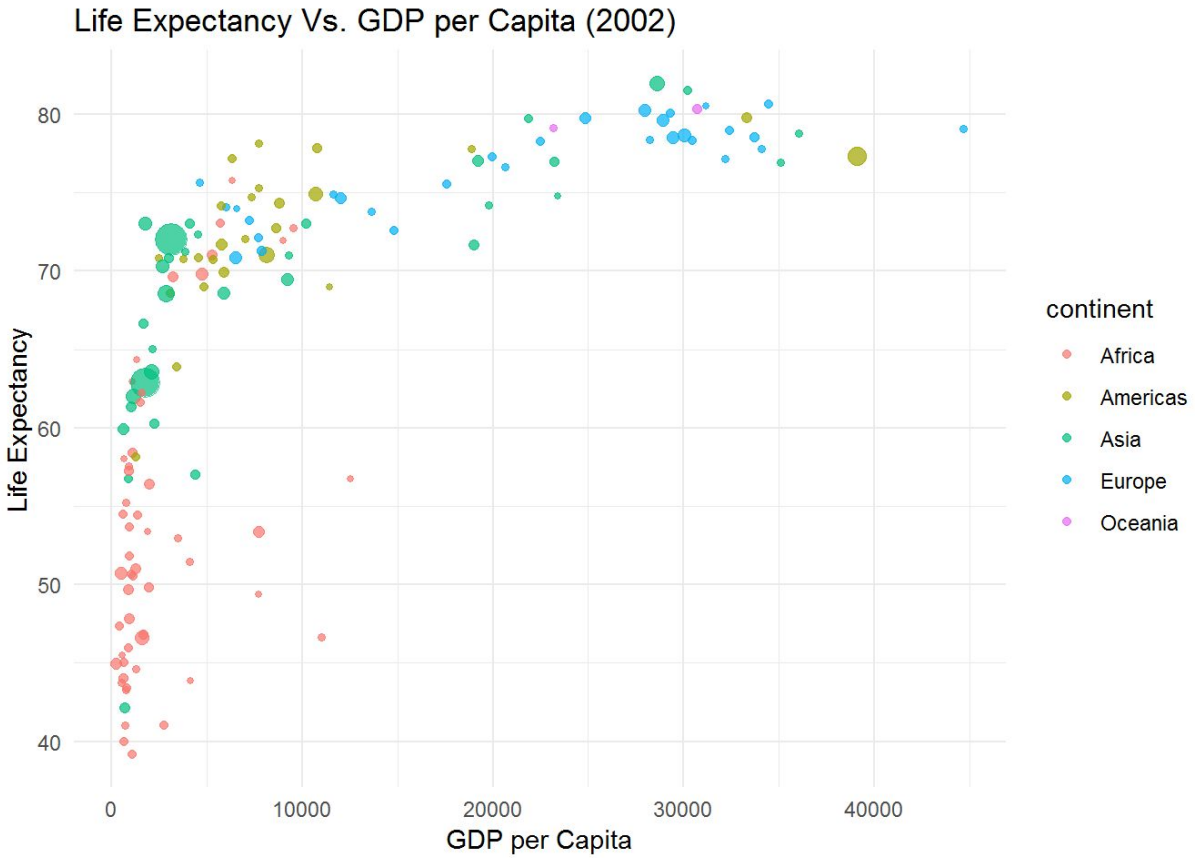
## PART 2

### googleplayappstore

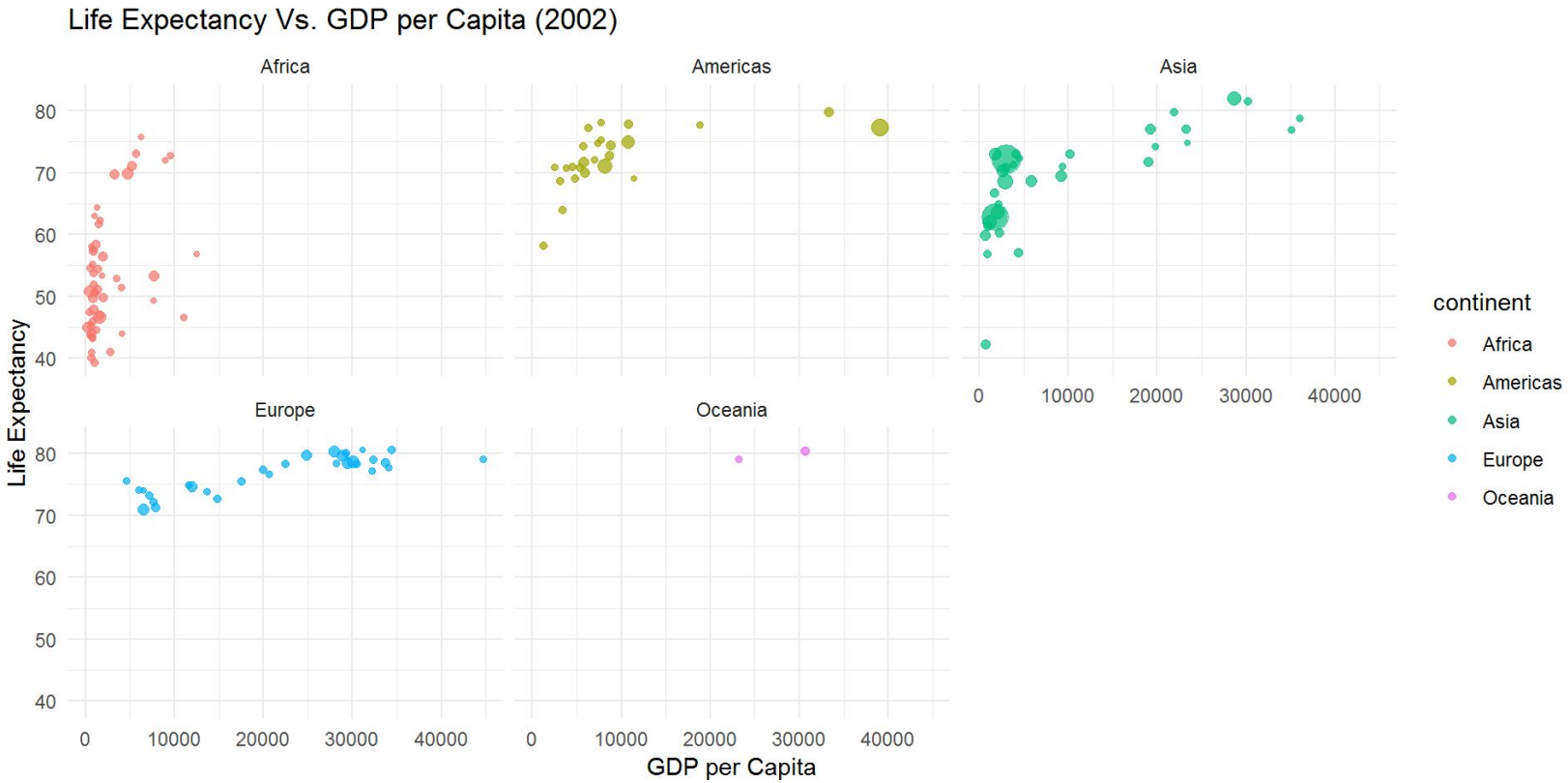
Main Source: Kaggle

https://www.kaggle.com/lava18/google-play-store-apps

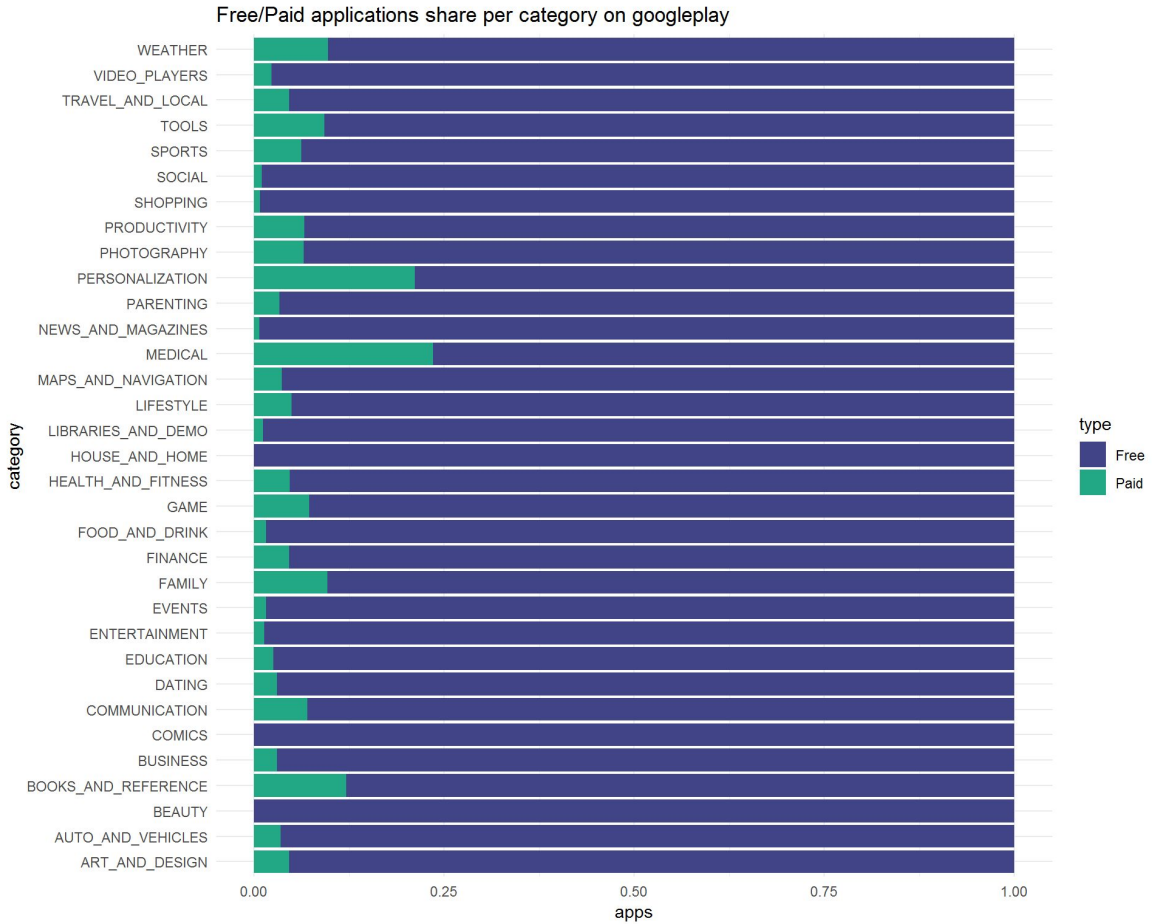
# What are we going to create today?



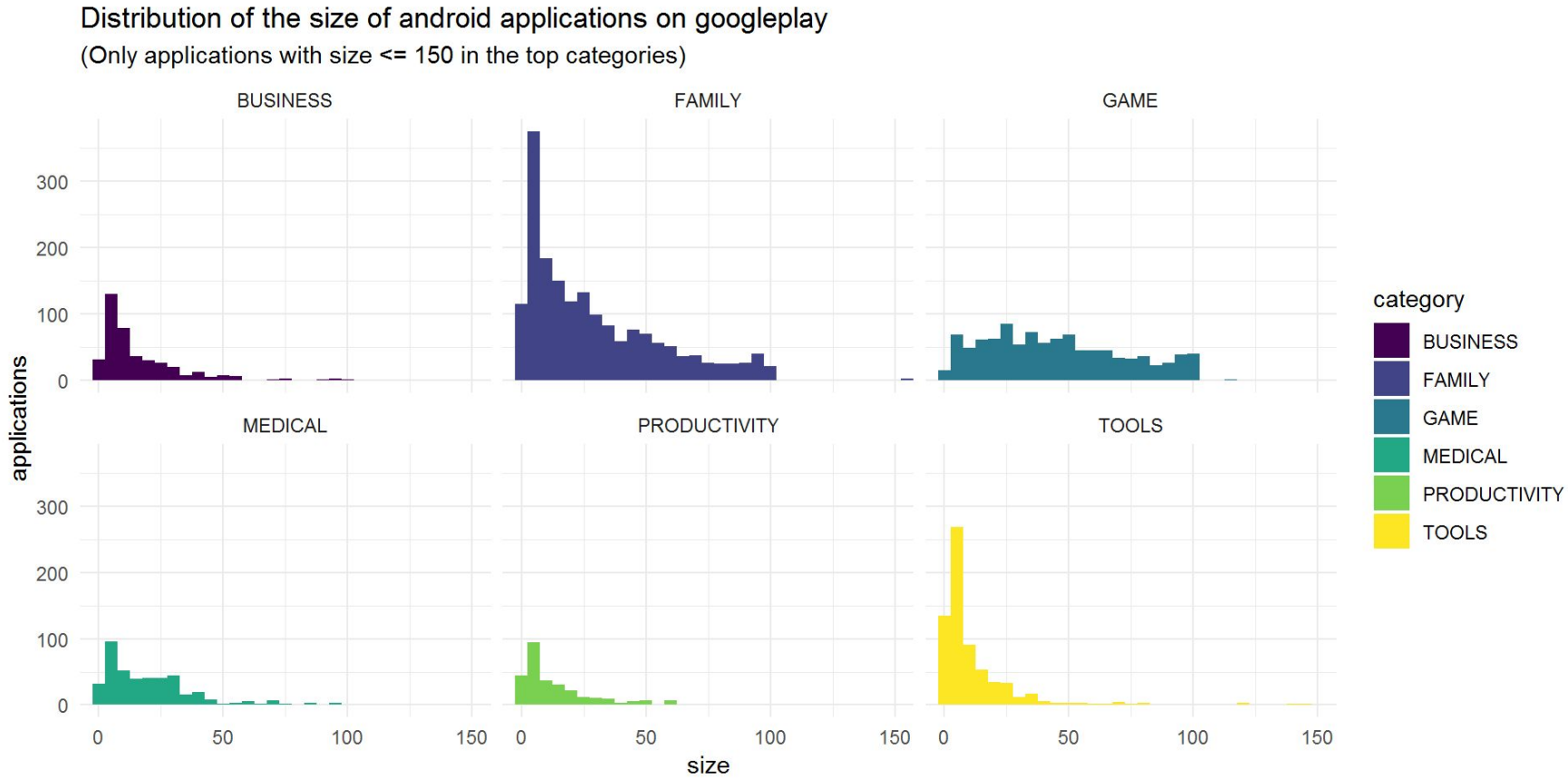
# What are we going to create today?



# What are we going to create today?



# What are we going to create today?



AND MORE...

# ggplot2 Layers/Building Blocks

THEME

COORDINATES

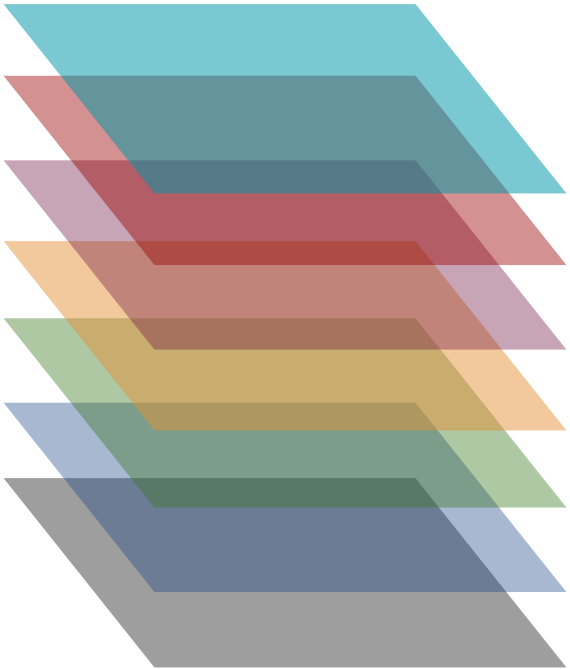
STATISTICS

FACETS

GEOMETRIES

AESTHETICS

DATA



# Gapminder Dataset

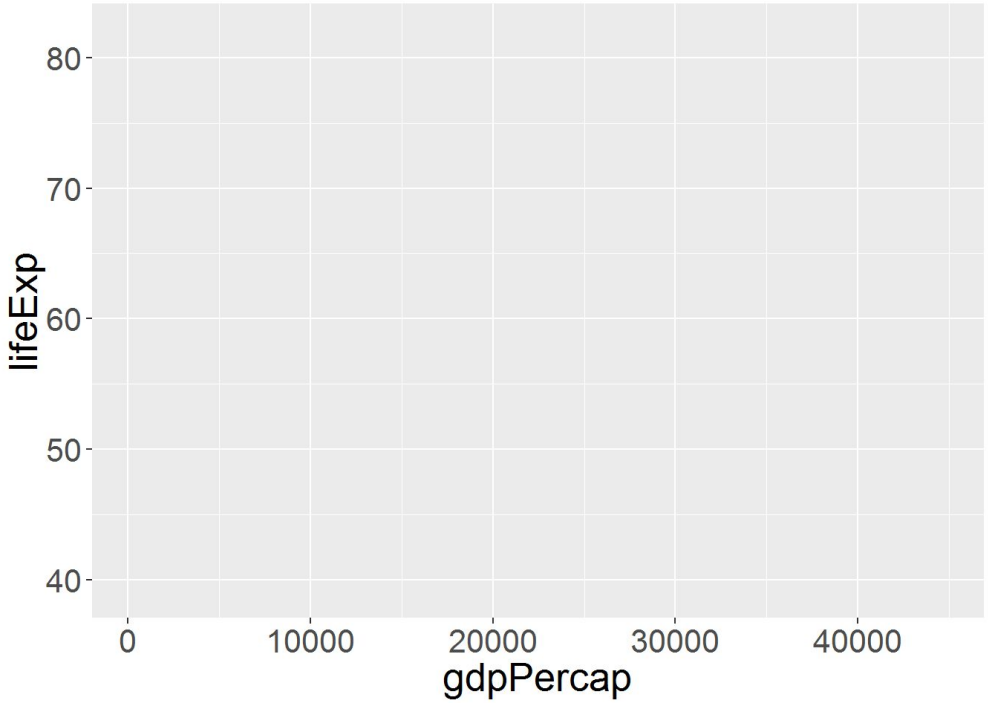
```
> library(gapminder)
> gapminder

# A tibble: 1,704 x 6
   country     continent  year lifeExp      pop gdpPercap
   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
 1 Afghanistan Asia       1952  28.801  8425333    779.45
 2 Afghanistan Asia       1957  30.332  9240934    820.85
 3 Afghanistan Asia       1962  31.997 10267083    853.10
 4 Afghanistan Asia       1967  34.02  11537966    836.20
 5 Afghanistan Asia       1972  36.088 13079460    739.98
 6 Afghanistan Asia       1977  38.438 14880372    786.11
 7 Afghanistan Asia       1982  39.854 12881816    978.01
 8 Afghanistan Asia       1987  40.822 13867957    852.40
 9 Afghanistan Asia       1992  41.674 16317921    649.34
10 Afghanistan Asia       1997  41.763 22227415    635.34
# ... with 1,694 more rows
```

DATA

AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp))
```



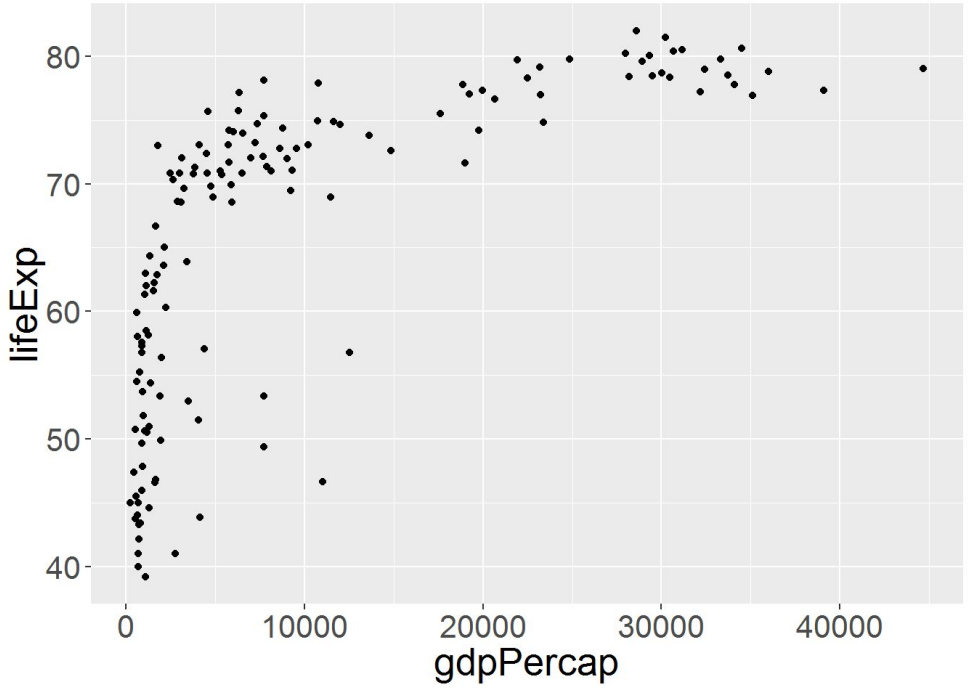
DATA

AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp))+
  geom_point()
```

GEOMETRY

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp))+
    geom_point()
```



DATA

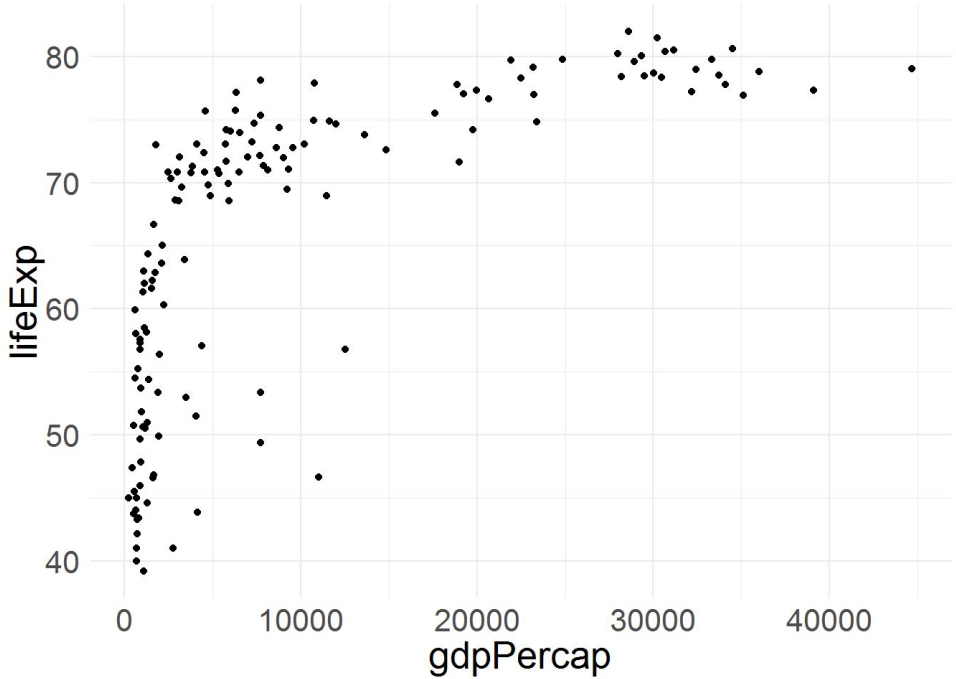
AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp))+
  geom_point()+
  theme_minimal()
```

GEOMETRY

THEME

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp))+
    geom_point()+
    theme_minimal()
```



DATA

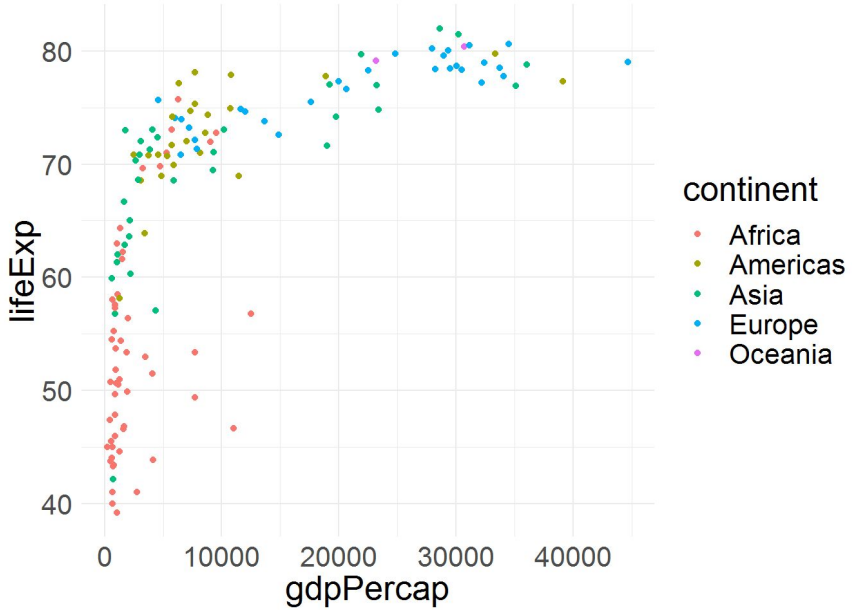
AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent))+
  geom_point()+
  theme_minimal()
```

GEOMETRY

THEME

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent))+
    geom_point()+
    theme_minimal()
```



DATA

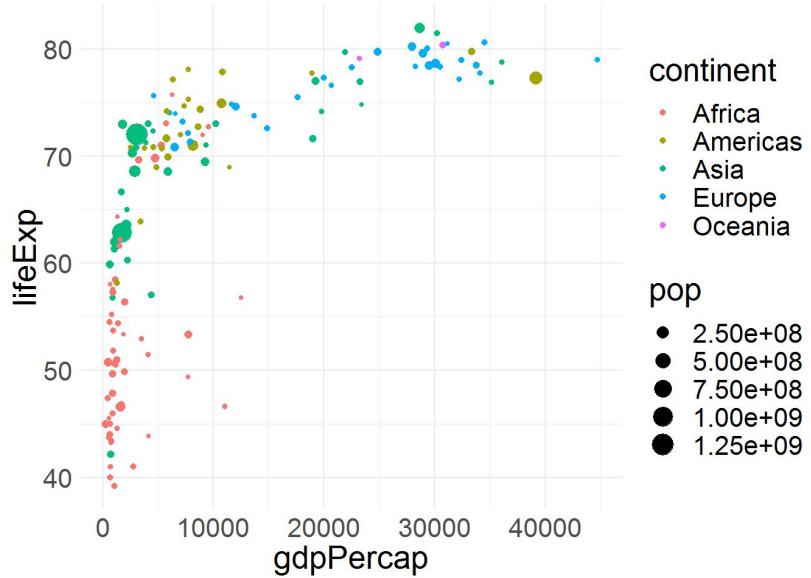
AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
  geom_point()+
  theme_minimal()
```

GEOMETRY

THEME

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
  geom_point()+
  theme_minimal()
```



DATA

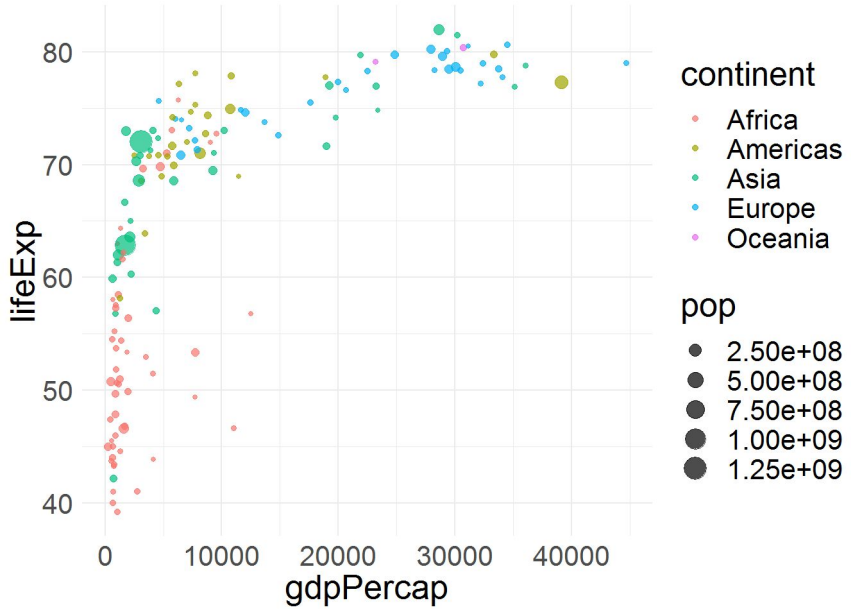
AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
  geom_point(alpha = 0.7)+
  theme_minimal()
```

GEOMETRY

THEME

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
    geom_point(alpha = 0.7)+
    theme_minimal()
```



DATA

AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
```

GEOMETRY

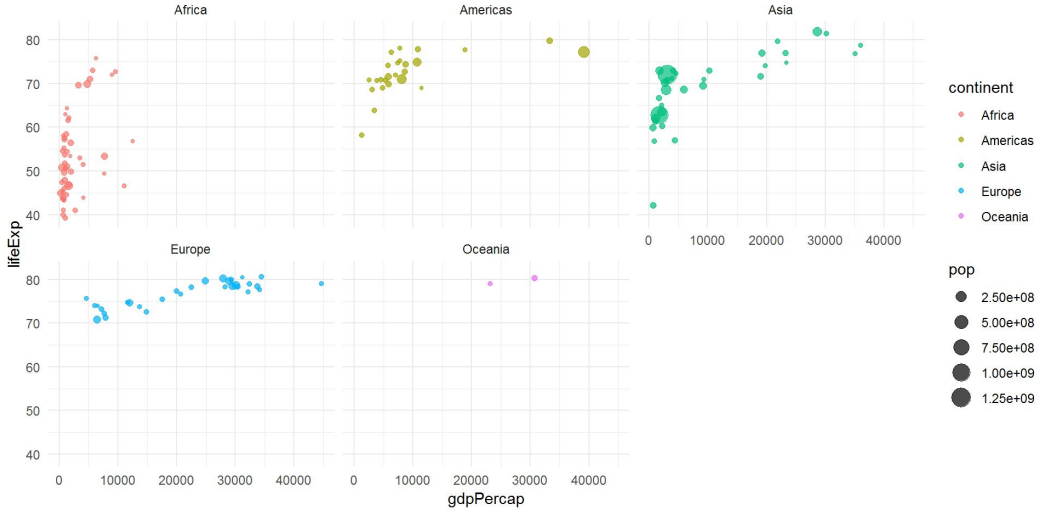
```
geom_point(alpha = 0.7)+
```

THEME

```
theme_minimal()+
facet_wrap(~continent)
```

FACETS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
  geom_point(alpha = 0.7)+
  theme_minimal()+
  facet_wrap(~continent)
```



DATA

AESTHETICS

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
```

GEOMETRY

```
geom_point(alpha = 0.7)+
```

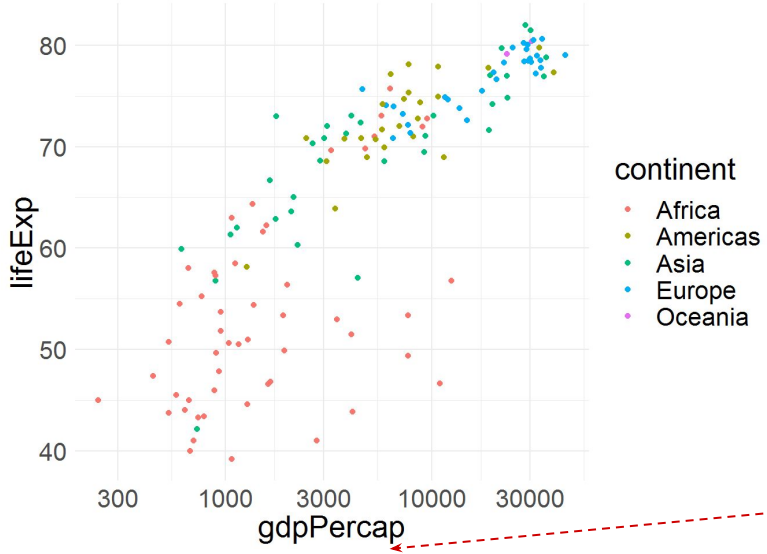
THEME

```
theme_minimal()+
```

```
scale_x_log10()
```

COORDINATES

```
ggplot(data = gapminder_2002, aes(x = gdpPercap, y = lifeExp,
                                  color = continent,
                                  size = pop))+
  geom_point(alpha = 0.7)+
  theme_minimal()+
  scale_x_log10()
```




NOTICE THE X-LABEL

# googleplaystore Dataset

```
> str(googleplaystore, max.level = 1)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':   10839 obs. of  13 variables:
 $ app           : chr  "Photo Editor & Candy Camera & Grid & ScrapBook" "Coloring book moana" "U
Launcher Lite – FREE Live Cool Themes, Hide Apps" "Sketch - Draw & Paint" ...
 $ category      : chr  "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" "ART_AND_DESIGN" ...
 $ rating        : num  4.1 3.9 4.7 4.5 4.3 4.4 3.8 4.1 4.4 4.7 ...
 $ reviews       : int  159 967 87510 215644 967 167 178 36815 13791 121 ...
 $ size          : num  19 14 8.7 25 2.8 5.6 19 29 33 3.1 ...
 $ installs      : chr  "10,000+" "500,000+" "5,000,000+" "50,000,000+" ...
 $ type          : chr  "Free" "Free" "Free" "Free" ...
 $ price         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ content_rating: chr  "Everyone" "Everyone" "Everyone" "Teen" ...
 $ genres        : chr  "Art & Design" "Art & Design;Pretend Play" "Art & Design" "Art & Design" ...
 $ last_updated  : Date, format: "2018-01-07" "2018-01-15" "2018-08-01" "2018-06-08" ...
 $ current_ver   : chr  "1.0.0" "2.0.0" "1.2.4" "Varies with device" ...
 $ android_ver   : chr  "4.0.3 and up" "4.0.3 and up" "4.0.3 and up" "4.2 and up" ...
 - attr(*, "spec")=List of 2
  ..- attr(*, "class")= chr "col_spec"
```

# googleplaystore Dataset

**Columns**

- A App Application name
- A Category Category the app belongs to ss
- # Rating Overall user rating of the app (as when scraped)
- # Reviews Number of user reviews for the app (as when scraped)
- A Size Size of the app (as when scraped)
- A Installs Number of user downloads/installs for the app (as when scraped)
- A Type Paid or Free
- A Price Price of the app (as when scraped)
- A Content Rating Age group the app is targeted at - Children / Mature 21+ / Adult
- A Genres An app can belong to multiple genres (apart from its main category). For eg, a musical family game will belong to Music, Game, Family genres.
- Last Updated Date when the app was last updated on Play Store (as when scraped)
- A Current Ver Current version of the app available on Play Store (as when scraped)
- A Android Ver Min required Android version (as when scraped)

https://www.kaggle.com/lava18/google-play-store-apps

# googleplaystore Dataset

- What is the total number of reviews for each category?
- What is the share of free/paid applications in each category?
- What is the distribution of applications size?

## CALCULATE AND PLOT