Bac France
A Story of Success Rates and Specialties

Introduction :

We are working with an official dataset containing information about the French baccalaureate results, broken down by academy, session, Path, and specialty. The main goal of this project is to learn how to use R Shiny for data management and visualization on a real-world dataset.

1) Why R ? :

a) Richness of statistical functions:

R natively provides all the statistical tools needed for educational data analysis (means, rates, correlations, etc.).

b) Specialized packages:

Packages like `dplyr` and `tidyr` make it easy to manipulate and reshape large datasets, while `ggplot2` and `plotly` allow for advanced, interactive visualizations.

c) Flexibility for complex datasets:

R can handle data from various sources (CSV, Excel, JSON) and is ideal for building a data transformation → analysis → visualization pipeline.

2) Data refinement :

a) Step 1: Load & Explore

- Read the CSV file and inspect the structure with str(df_working) and summary(df_working).

- Check for missing values and data consistency.

b) Step 2: Filter & Prepare

- Filter by session, academy, or specialty as needed for focused analysis.
- Example:

```
general <- df_working %>% filter(Path == "Général")
```

c) Step 3: Transform for Visualization

- Aggregate data by Path, Academy, or Session for summary statistics.
- Example:

```
summary_by_path <- df_working %>%
   group_by(Path) %>%
   summarise(
     Admitted = sum(Total_Number_Admitted, na.rm = TRUE),
     Attendees = sum(Number_of_Attendees, na.rm = TRUE),
     SuccessRate = round(Admitted / Attendees * 100, 2)
   )
```

d) Step 4: Enrich with statistics

- Calculate overall and per-path success rates, number of mentions, and rejected candidates.
- Example:

```
stats_summary <- df_working %>%
    group_by(Path) %>%
    summarise(
      mean_success = mean(Total_Number_Admitted /
Number_of_Attendees, na.rm = TRUE),
      sd_success = sd(Total_Number_Admitted / Number_of_Attendees,
na.rm = TRUE)
    )
```

3) **Data analysis with statistical techniques :**

<u>a) Central Tendencies:</u>

- The overall mean success rate is high (often above 85%), but varies by Path and Academy.
- The Général Path typically has the highest success rate, followed by Technologique and Professionnel.

<u>b) Dispersion Measures:</u>

- The standard deviation of success rates is higher in the Professionnel Path, indicating more variability between academies or sessions.
- The Général Path shows the most stable results.

4) <u>Data analysis with visualization :</u>

<u>a) Barplots & Pie Charts:</u>

- Barplots show the number of admitted students per Path or Academy.
- Pie charts (camemberts) visualize the proportion of each Path in total admissions or success rates.

<u>b) Boxplots:</u>

- Boxplots by Academy or Session reveal outliers and the spread of success rates.

<u>c) Time Series:</u>

- Line charts can be used to track the evolution of success rates over multiple sessions.

<u>d) Interactive Dashboards:</u>

- Filters allow dynamic exploration by year, academy, or specialty.
- KPIs (Key Performance Indicators) display overall success rates, number of admitted, and number of rejected candidates.

5) <u>Key insights :</u>

<u>a) Général Path – The Benchmark:</u>

Consistently the highest success rate and the largest number of candidates. Low variability across academies.

<u>b) Technologique Path – The Middle Ground:</u>

Intermediate success rates, with some academies performing close to the Général Path.

<u>c) Professionnel Path – The Challenge:</u>

Lower and **more variable** success rates.
Some academies show significant **improvement** or **decline** over time.

<u>d) Mentions and Rejected:</u>

The distribution of **honors** is skewed towards the **Général** Path.
The number of **rejected** candidates is **highest** in the **Professionnel** Path.

<u>e) Temporal Trends:</u>

Success rates have remained stable or improved slightly in recent years.

6) <u>Conclusion:</u>

The bac dataset lets us explore and visualize the status of education in France. R, with its ecosystem of packages, enables efficient data cleaning, transformation, and insightful reporting on this kind of dataset.