

# Education quality

Daphné PERRY Justine CATTA Chloé MARQUIS Clémence HUANG

## Table of contents

Baccalaureate success rate . . . . .	1
Access rate to the baccalaureate . . . . .	2
Baccalaureate honors (mentions) rate . . . . .	2
Estimation of expected rates . . . . .	3
Graphical representations . . . . .	6
Descriptive analyses . . . . .	6
Correlation with IPS . . . . .	7

1.

What are the different high school quality indicators published by the ministry of higher education? <sup>1</sup>

The French Ministry of Education, through the DEPP (Direction de l'évaluation, de la prospective et de la performance), publishes annually the *High School Value-Added Indicators* (Indicateurs de Valeur Ajoutée des Lycées, IVAL). Their objective is to assess high school performance while accounting for differences in student intake, rather than relying solely on raw outcome measures.

Three main **quality indicators** are published for general and technological high schools:

### Baccalaureate success rate

This indicator measures the proportion of students present at the baccalaureate examination who obtain the diploma:

$$\text{Success rate} = \frac{\text{Number of students who pass the bac}}{\text{Number of students present at the exam}}$$

---

<sup>1</sup><https://www.education.gouv.fr/les-indicateurs-de-resultats-des-colleges-et-des-lycees-377729>  
[https://data.education.gouv.fr/explore/dataset/fr-en-indicateurs-de-resultat-des-lycees-gt\\_v2/](https://data.education.gouv.fr/explore/dataset/fr-en-indicateurs-de-resultat-des-lycees-gt_v2/)

## Access rate to the baccalaureate

The access rate represents the probability that a student entering a given high school in *seconde* will eventually obtain the baccalaureate in the same establishment, even if the student repeats a grade. It is computed as the product of successive transition probabilities:

$$\text{Access rate}_{2\text{nde} \rightarrow \text{bac}} = (2\text{nde} \rightarrow 1\text{\`ere}) \times (1\text{\`ere} \rightarrow \text{Terminale}) \times (\text{Terminale} \rightarrow \text{Bac})$$

## Baccalaureate honors (mentions) rate

This indicator measures the proportion of students present at the baccalaureate examination who obtain a mention (Assez Bien, Bien, or Très Bien, with or without Félicitations).

For most indicators, results are published both **overall** and **by track or series** (general and technological streams such as STI2D, STMG, STL, ST2S, etc.). Each indicator is reported as an **observed value**, an **expected value**, and a **value-added measure**.

Administrative variables available in the datasets (UAI code, location, sector, student counts by grade) are descriptive and do not constitute quality indicators.

2.

How is the value-added measure (IVAL, indicateur de valeur ajoutée), calculated? <sup>2</sup>

The **value-added indicator** measures the contribution of a high school to student outcomes, net of external factors. It is defined as the difference between the observed rate and the expected rate:

$$\text{Value-added} = \text{Observed rate} - \text{Expected rate}$$

The expected rate is the average rate of high schools with students that have identical characteristics. Therefore the IVAL evaluates the contribution due to the establishment, given initial students' profiles. A positive value-added indicates that on average the high school allows students to perform better than expected given their initial profile, while a negative value-added indicates underperformance.

---

<sup>2</sup><https://catalogue.depp.education.fr/index.php/catalog/274>

Franck Evain (Insee), *Indicateurs de valeur ajoutée des lycées*: <https://www.insee.fr/fr/information/5008703?sommaire=5008710>

## Estimation of expected rates

Expected rates are estimated using logistic models (often multilevel) for binary outcomes (success or failure, mention or no mention) and account for the hierarchical structure of the data (students nested within schools).

The models include:

**Individual-level variables** - Age - Gender - Academic achievement at entry into high school (e.g., results at the Diplôme National du Brevet) - Social background, measured using the Social Position Index (IPS)

**School-level variables** - Proportion of students with academic delay - Proportion of girls - Average IPS of students - Average prior academic achievement within the school

For each student, the model produces an expected probability of success, access, or obtaining a mention, assuming enrollment in a statistically average high school (i.e., excluding the specific school effect). The expected rate of a high school is then computed as the average of its students' expected probabilities.

In the case of access to the baccalaureate, separate models are estimated for each transition (seconde→première, première→terminale, terminale→bac), and the expected access rate is obtained by multiplying the expected transition probabilities.

3.

How do the IVAL and other indicators correlate with IPS (a higher IPS indicates a higher average socio-economic status of parents at the school level) in Paris? Proposed graphical representations are shown separately for public and private high schools.

```
library(vroom)
library(here)
library(ggplot2)
library(dplyr)
library(readxl)
library(tidyr)
library(stringr)
library(janitor)

here::i_am("quality_education.Rproj")

# Import datasets
quality_indicators <- vroom(here("data", "fr-en-indicateurs-de-resultat-des-lycees-gt_v2.csv"),
  clean_names()
```

```

ips <- vroom(here("data", "fr-en-ips_lycees.csv")) |>
  clean_names()

path_ival <- here("data", "ival-2024-154664.xlsx")
ival_lgt <- read_excel(path_ival, sheet = "LGT", skip = 1) |>
  clean_names()
ival_lp <- read_excel(path_ival, sheet = "LP", skip = 1) |>
  clean_names()

# Convert logical columns (mostly NA) to numeric
quality_indicators <- quality_indicators |>
  mutate(across(where(is.logical), ~ as.numeric(.)))

# Build year from school-year label (robust to possible suffixes)
rentree_col <- names(ips)[str_detect(names(ips), "rentree")][1]
if (is.na(rentree_col)) stop("No column containing 'rentree' found in ips.")

ips <- ips |>
  mutate(annee = as.integer(str_extract(as.character(.data[[rentree_col]]), "\\d{4}")))

# Keep Paris (dept 75)
quality_paris <- quality_indicators |>
  filter(code_departement == "75")

ips_paris <- ips |>
  filter(code_du_departement == "75")

# Pick the most recent common year
years_common <- intersect(sort(unique(quality_paris$annee)), sort(unique(ips_paris$annee)))
if (length(years_common) == 0) stop("No common year in Paris (75).")
year_use <- max(years_common, na.rm = TRUE)

quality_paris_y <- quality_paris |>
  filter(annee == year_use)

ips_paris_y <- ips_paris |>
  filter(annee == year_use)

# Harmonize sector labels
quality_paris_y <- quality_paris_y |>
  mutate(
    secteur = case_when(

```

```

      str_detect(tolower(secteur), "public") ~ "Public",
      str_detect(tolower(secteur), "priv") ~ "Privé",
      TRUE ~ as.character(secteur)
    )
  )

ips_paris_y <- ips_paris_y |>
  mutate(
    secteur = case_when(
      str_detect(tolower(secteur), "public") ~ "Public",
      str_detect(tolower(secteur), "priv") ~ "Privé",
      TRUE ~ as.character(secteur)
    )
  )

# Merge (UAI + year + sector)
df_paris <- quality_paris_y |>
  left_join(
    ips_paris_y |>
      select(
        uai, annee, secteur,
        matches("^ips_"),
        matches("^ecart_type_de_l_ips")
      ),
    by = c("uai", "annee", "secteur")
  )

# Quick join check
df_paris |>
  summarise(
    n_qi = n(),
    n_ips_match = sum(!is.na(ips_ensemble_gt_pro)),
    pct_ips_match = mean(!is.na(ips_ensemble_gt_pro)) * 100
  ) |>
  print()

```

```

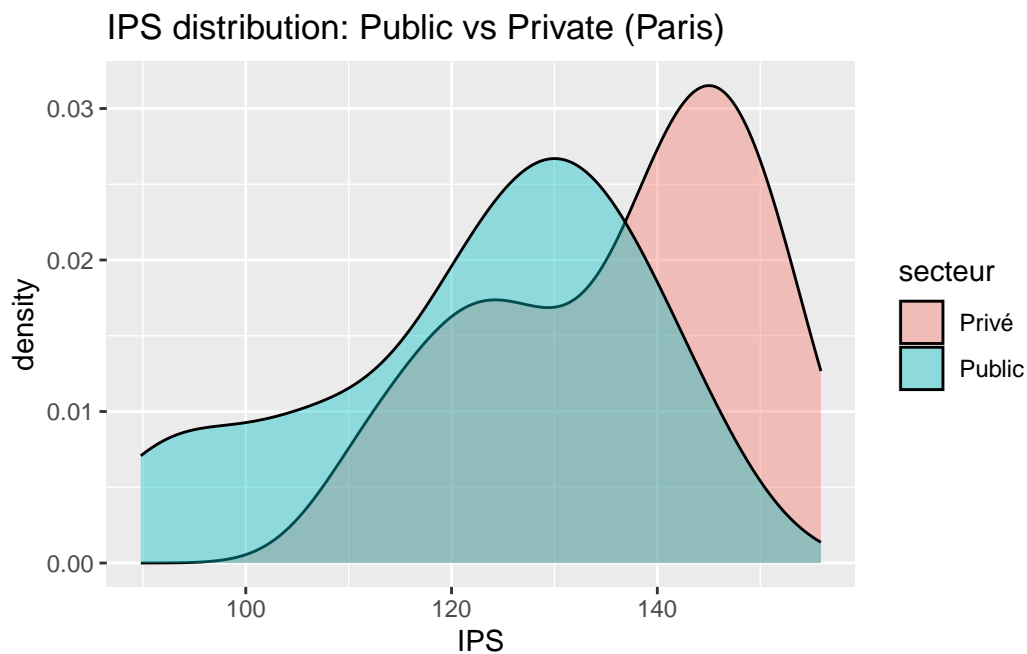
# A tibble: 1 x 3
  n_qi n_ips_match pct_ips_match
  <int>    <int>         <dbl>
1   107        107          100

```

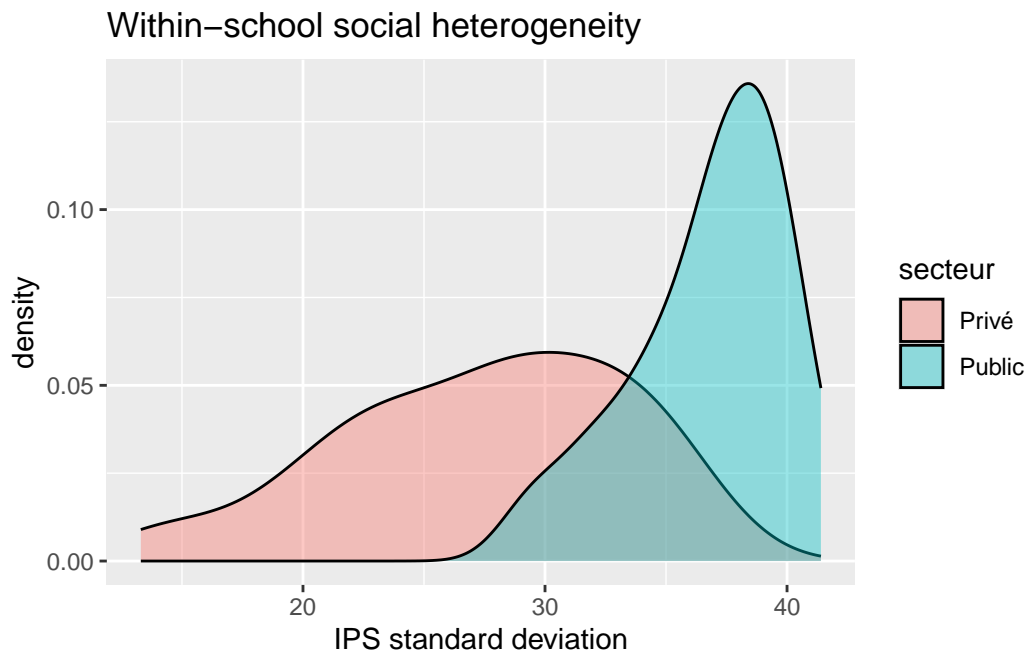
## Graphical representations

### Descriptive analyses

```
ggplot(df_paris, aes(x = ips_ensemble_gt_pro, fill = secteur)) +  
  geom_density(alpha = 0.4) +  
  labs(  
    x = "IPS",  
    title = "IPS distribution: Public vs Private (Paris)"  
  )
```



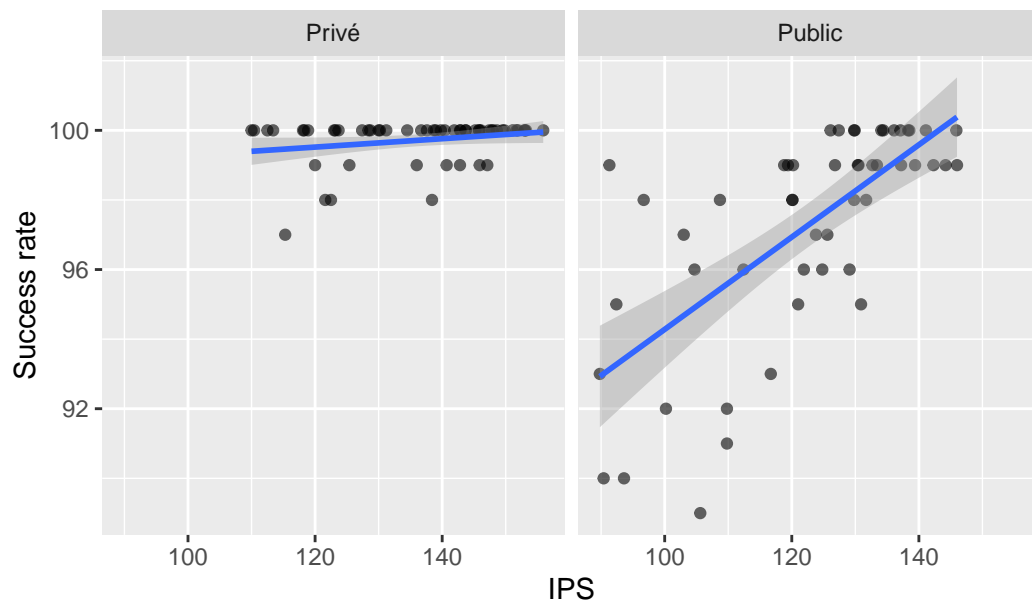
```
if ("ecart_type_de_l_ips_voie_gt" %in% names(df_paris)) {  
  ggplot(df_paris, aes(x = ecart_type_de_l_ips_voie_gt, fill = secteur)) +  
    geom_density(alpha = 0.4) +  
    labs(  
      x = "IPS standard deviation",  
      title = "Within-school social heterogeneity"  
    )  
} else {  
  cat("Variable ecart_type_de_l_ips_voie_gt not found in df_paris.\n")  
}
```



### Correlation with IPS

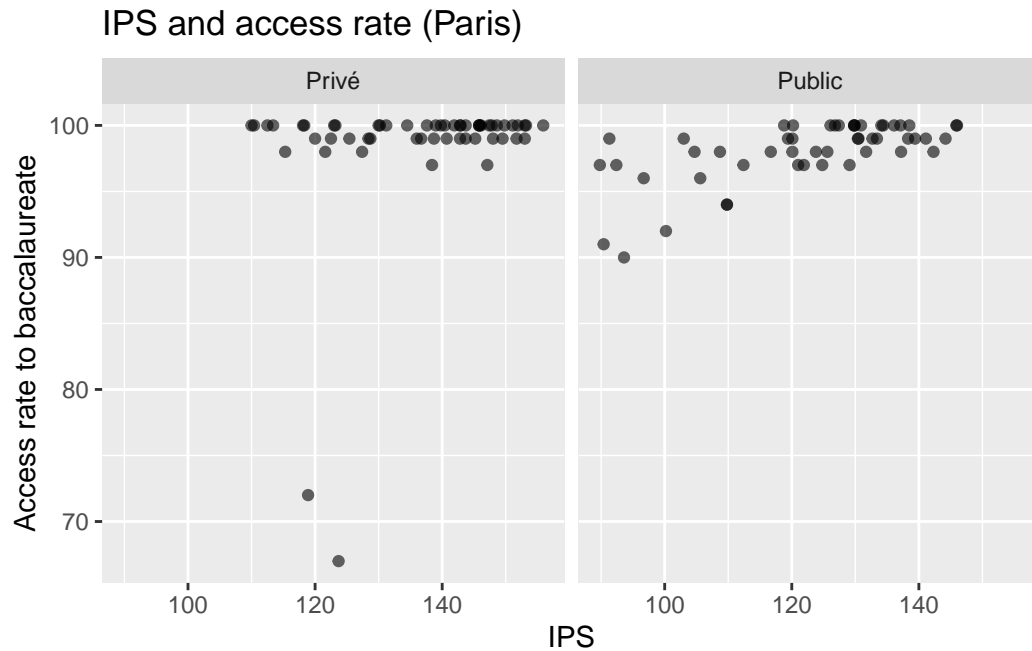
```
ggplot(df_paris,  
  aes(x = ips_ensemble_gt_pro,  
    y = taux_de_reussite_toutes_series)) +  
  geom_point(alpha = 0.6) +  
  geom_smooth(method = "lm", se = TRUE) +  
  facet_wrap(~ secteur) +  
  labs(  
    x = "IPS",  
    y = "Success rate",  
    title = "IPS and success rate (Paris)"  
  )
```

IPS and success rate (Paris)



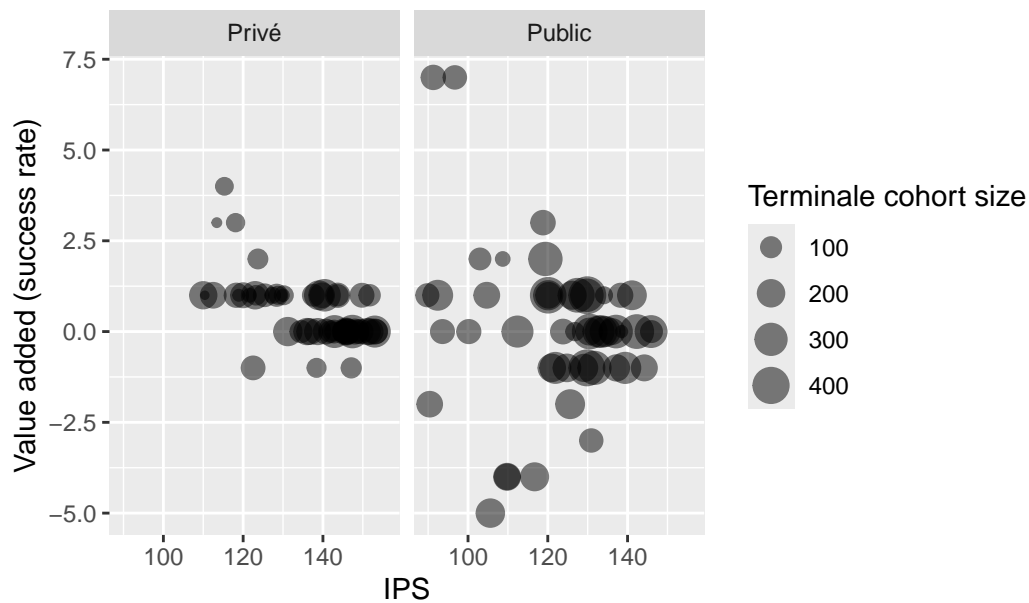
```
ggplot(df_paris,
       aes(x = ips_ensemble_gt_pro,
           y = taux_daces_terminale_bac)) +
  geom_point(alpha = 0.6) +
  facet_wrap(~ secteur) +
  labs(
    x = "IPS",
    y = "Access rate to baccalaureate",
    title = "IPS and access rate (Paris)"
  )
```



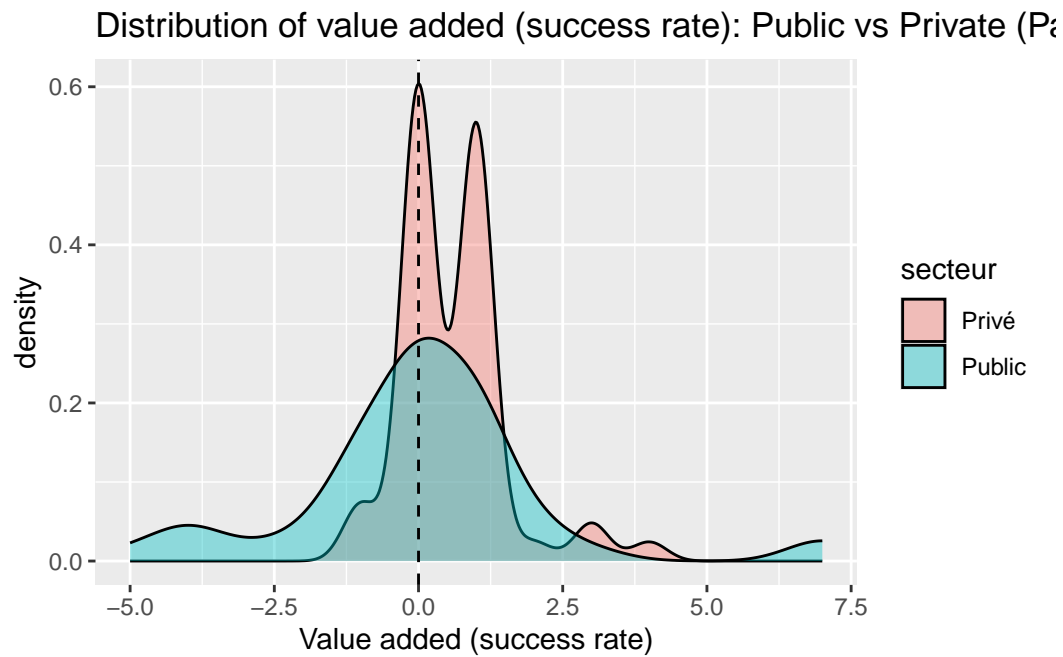


```
ggplot(df_paris,
  aes(x = ips_ensemble_gt_pro,
      y = valeur_ajoutee_du_taux_de_reussite_toutes_series,
      size = effectif_de_terminale)) +
  geom_point(alpha = 0.5) +
  facet_wrap(~ secteur) +
  labs(
    x = "IPS",
    y = "Value added (success rate)",
    size = "Terminale cohort size",
    title = "IPS and value added (success rate) with cohort size (Paris)"
  )
```

IPS and value added (success rate) with cohort size (Paris)

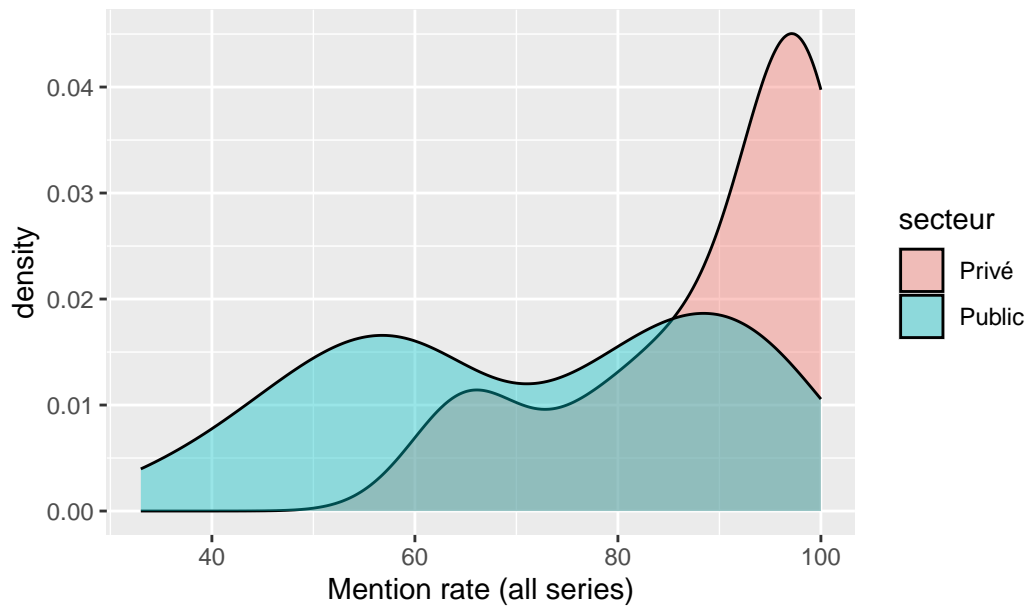


```
ggplot(df_paris,
  aes(x = valeur_ajoutee_du_taux_de_reussite_toutes_series,
    fill = secteur)) +
  geom_density(alpha = 0.4) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(
    x = "Value added (success rate)",
    title = "Distribution of value added (success rate): Public vs Private (Paris)"
  )
```



```
ggplot(df_paris,  
  aes(x = taux_de_mentions_toutes_series, fill = secteur)) +  
  geom_density(alpha = 0.4) +  
  labs(  
    x = "Mention rate (all series)",  
    title = "Distribution of mention rates: Public vs Private (Paris)"  
  )
```

Distribution of mention rates: Public vs Private (Paris)



```
ggplot(df_paris,  
  aes(x = valeur_ajoutee_du_taux_de_mentions_toutes_series,  
      fill = secteur)) +  
geom_density(alpha = 0.4) +  
geom_vline(xintercept = 0, linetype = "dashed") +  
labs(  
  x = "Value added (mention rate)",  
  title = "Distribution of value added (mention rate): Public vs Private (Paris)"  
)
```

