

Projet de programmation statistique avec R

Leo Jean UNITE; Diego CASAS BARCENAS

2024-12-17

```
knitr::opts_chunk$set(echo = TRUE, warning = F, message = F)
```

```
source("Fonction.R")
source("Data.R")
```

```
## # A tibble: 17 × 5
##   move1w num_games white_wins black_wins white_odds
##   <chr>      <int>      <dbl>      <dbl>      <dbl>
## 1 g3         18777       7543.       5323.       1.42
## 2 d4        1308961     504870.     370724.     1.36
## 3 c4         259049      97642.     73897.     1.32
## 4 Nf3        142996      53966.     41511.     1.30
## 5 e4        1691765     650893.     536512.     1.21
## 6 b3         16134       6051.       5527.       1.09
## 7 00E+04       7040      2746.       2640        1.04
## 8 Nc3         4480       1659.       1647.       1.01
## 9 a3          1308        467.        506.       0.922
## 10 b4         6215      2309.       2504.       0.922
## 11 h3          183        67.0        75.9       0.882
## 12 d3         1247        443.        502.       0.882
## 13 f4        16315      5660.     6554.       0.864
## 14 c3          799        258.        330.       0.782
## 15 g4          797        279.        371.       0.752
## 16 h4          121        41.0        57.0       0.720
## 17 00E+03      1713       553.       805.       0.687
```

Objectifs

L'objectif de ce projet est d'appliquer les connaissances acquises pendant le cours de programmation statistique avec R. Chaque groupe doit s'approprier sa base de données, en faire des résumés statistiques et réaliser différents graphiques d'analyse descriptive. Les graphiques et les statistiques diffèrent en fonction des sujets, il faut choisir lesquels sont pertinents à réaliser.

Déroulement

Le projet a lieu sur 3 séances de travail : le 17/12, le 18/12 et le 07/01. Le rendu final est attendu le 12/01 à 22h au plus tard. Les présentations orales auront lieu le 15/01 après-midi. Les modalités précises de l'oral seront communiquées ultérieurement, mais celui-ci portera principalement sur vos choix méthodologiques, vos analyses et votre code R. Cahier des charges Votre projet devra inclure au minimum les éléments suivants :

1. Gestion des données

- Importation des données dans R.

- Description générale des données (par exemple : nombre de lignes, colonnes, et valeurs manquantes).

```
dim(data)
```

```
## [1] 1884 24
```

2. Analyse descriptive

- Résumés statistiques pertinents pour chaque variable : moyennes, médianes, écarts types, effectifs, etc.
- Résumés croisés entre au moins deux variables.
- Analyses par sous-groupes ou populations spécifiques.

```
Analyse <- data %>% summary  
Analyse
```

```

## opening_name      side      num_games      ECO
## Length:1884      Length:1884      Min.   : 100.0      Length:1884
## Class :character  Class :character  1st Qu.: 314.8      Class :character
## Mode  :character  Mode  :character  Median : 788.5      Mode  :character
##                                     Mean   : 1846.0
##                                     3rd Qu.: 2225.0
##                                     Max.   :22482.0
## last_played_date  perf_rating  avg_player  perc_player_win
## Length:1884      Min.   :1583      Min.   :1577      Min.   : 7.50
## Class :character  1st Qu.:2157      1st Qu.:2166      1st Qu.:28.90
## Mode  :character  Median :2252      Median :2255      Median :35.10
##                                     Mean   :2236      Mean   :2237      Mean   :35.16
##                                     3rd Qu.:2329      3rd Qu.:2326      3rd Qu.:41.12
##                                     Max.   :2536      Max.   :2492      Max.   :77.60
## perc_draw         pec_opponent_win  moves_list      move1w
## Min.   : 4.00      Min.   : 6.70      Length:1884      Length:1884
## 1st Qu.:24.50      1st Qu.:28.90      Class :character  Class :character
## Median :29.40      Median :34.65      Mode  :character  Mode  :character
## Mean   :29.91      Mean   :34.93
## 3rd Qu.:34.62      3rd Qu.:40.52
## Max.   :68.50      Max.   :77.50
## move1b            move2w            move2b            move3w
## Length:1884      Length:1884      Length:1884      Length:1884
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## move3b            move4w            move4b            perc_white_win
## Length:1884      Length:1884      Length:1884      Min.   :13.60
## Class :character  Class :character  Class :character  1st Qu.:34.80
## Mode  :character  Mode  :character  Mode  :character  Median :39.10
##                                     Mean   :39.75
##                                     3rd Qu.:44.00
##                                     Max.   :77.60
## perc_black_win    white_odds        white_wins        black_wins
## Min.   : 6.70      Min.   :0.3086      Min.   : 21.0      Min.   : 8.946
## 1st Qu.:25.10      1st Qu.:1.0373      1st Qu.: 124.7      1st Qu.: 91.026
## Median :29.90      Median :1.3250      Median : 310.3      Median : 230.952
## Mean   :30.34      Mean   :1.4487      Mean   : 708.8      Mean   : 557.052
## 3rd Qu.:35.00      3rd Qu.:1.6741      3rd Qu.: 824.2      3rd Qu.: 651.263
## Max.   :64.80      Max.   :9.8101      Max.   :8295.9      Max.   :8700.534

```

```

# AnalyseCroisser
res

```

```
## # A tibble: 17 × 5
##   move1w num_games white_wins black_wins white_odds
##   <chr>      <int>      <dbl>      <dbl>      <dbl>
## 1 g3         18777      7543.      5323.      1.42
## 2 d4        1308961     504870.     370724.     1.36
## 3 c4         259049      97642.     73897.     1.32
## 4 Nf3        142996      53966.     41511.     1.30
## 5 e4        1691765     650893.     536512.     1.21
## 6 b3         16134       6051.     5527.     1.09
## 7 00E+04       7040       2746.     2640.     1.04
## 8 Nc3         4480       1659.     1647.     1.01
## 9 a3          1308        467.      506.     0.922
## 10 b4         6215       2309.     2504.     0.922
## 11 h3          183        67.0      75.9     0.882
## 12 d3         1247       443.      502.     0.882
## 13 f4        16315      5660.     6554.     0.864
## 14 c3          799       258.      330.     0.782
## 15 g4          797       279.      371.     0.752
## 16 h4          121       41.0      57.0     0.720
## 17 00E+03      1713       553.      805.     0.687
```

AnalyseGroupes

```
## # A tibble: 892 × 24
## # Groups:   side [1]
##   opening_name side num_games ECO last_played_date perf_rating avg_player
##   <chr>      <chr>      <int> <chr> <chr>          <int>      <int>
## 1 Alekhine Defen... white      692 B03 2018-06-22      2247      2225
## 2 Alekhine Defen... white     6485 B03 2018-07-06      2244      2194
## 3 Alekhine Defen... white      881 B03 2018-06-20      2187      2130
## 4 Alekhine Defen... white     1207 B03 2018-07-01      2243      2190
## 5 Alekhine Defen... white      334 B02 2018-06-27      2205      2178
## 6 Alekhine Defen... white      931 B02 2018-07-09      2174      2174
## 7 Alekhine Defen... white      501 B04 2018-06-25      2397      2336
## 8 Alekhine Defen... white      120 B04 2018-06-11      2097      2081
## 9 Alekhine Defen... white      269 B04 2018-03-04      2442      2383
## 10 Alekhine Defen... white      183 B05 2018-05-31      2305      2269
## # i 882 more rows
## # i 17 more variables: perc_player_win <dbl>, perc_draw <dbl>,
## #   pec_opponent_win <dbl>, moves_list <chr>, move1w <chr>, move1b <chr>,
## #   move2w <chr>, move2b <chr>, move3w <chr>, move3b <chr>, move4w <chr>,
## #   move4b <chr>, perc_white_win <dbl>, perc_black_win <dbl>, white_odds <dbl>,
## #   white_wins <dbl>, black_wins <dbl>
```

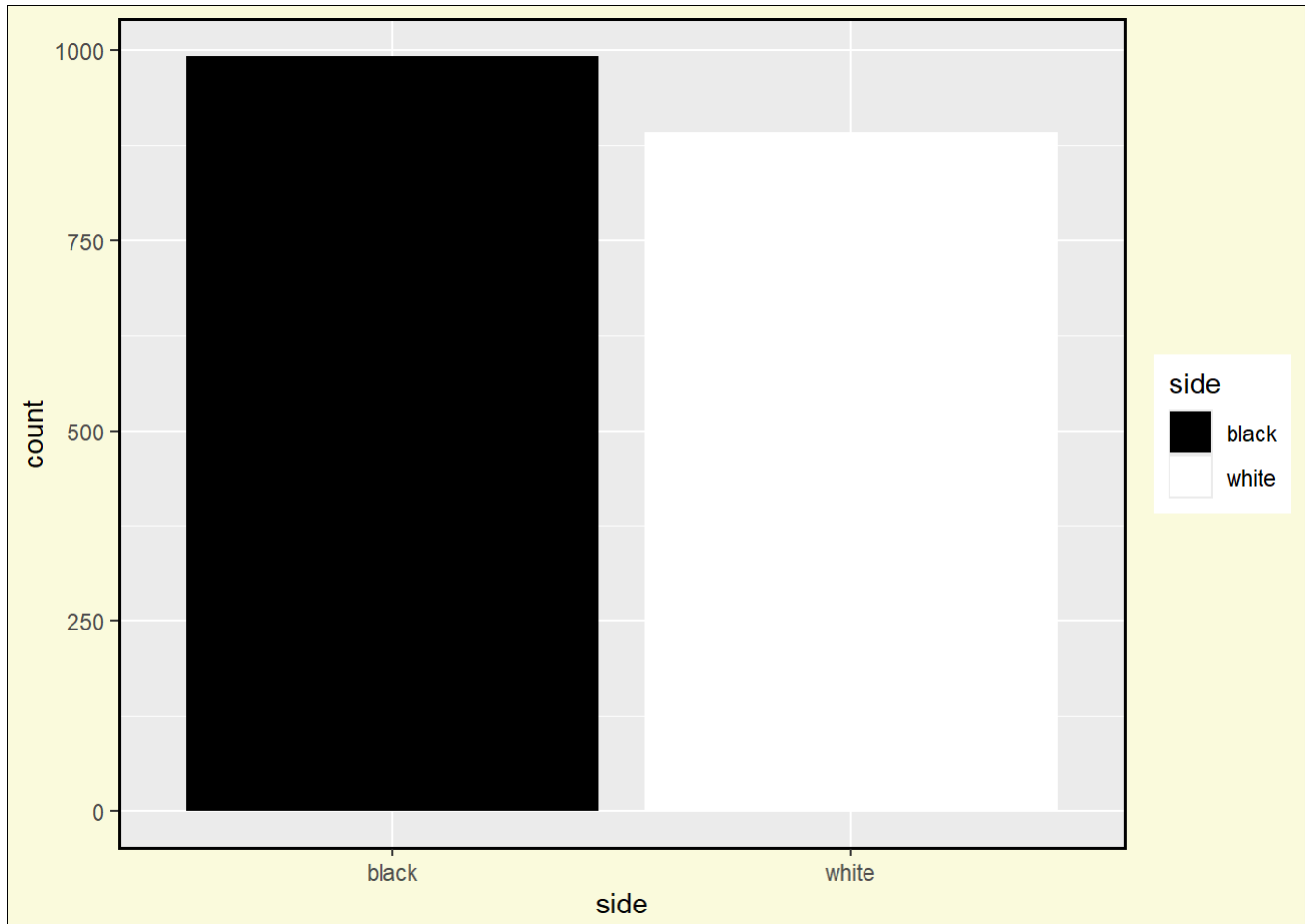
3. Visualisations graphiques

- Graphiques descriptifs pour des variables individuelles.

```
Pindi1 <- ggplot(data, aes(x = side))
Pindi1 <- Pindi1 + geom_bar(aes(fill = side)) + scale_fill_manual(values = c("black" = "black", "white" = "white")) + theme(plot.background = element_rect(fill = "#fdfcdc", color = "black"), # Changement de couleur de fond et ajouts des bordures
                        panel.border = element_rect(color = "black", fill = NA, size = 1))
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.  
## i Please use the `linewidth` argument instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

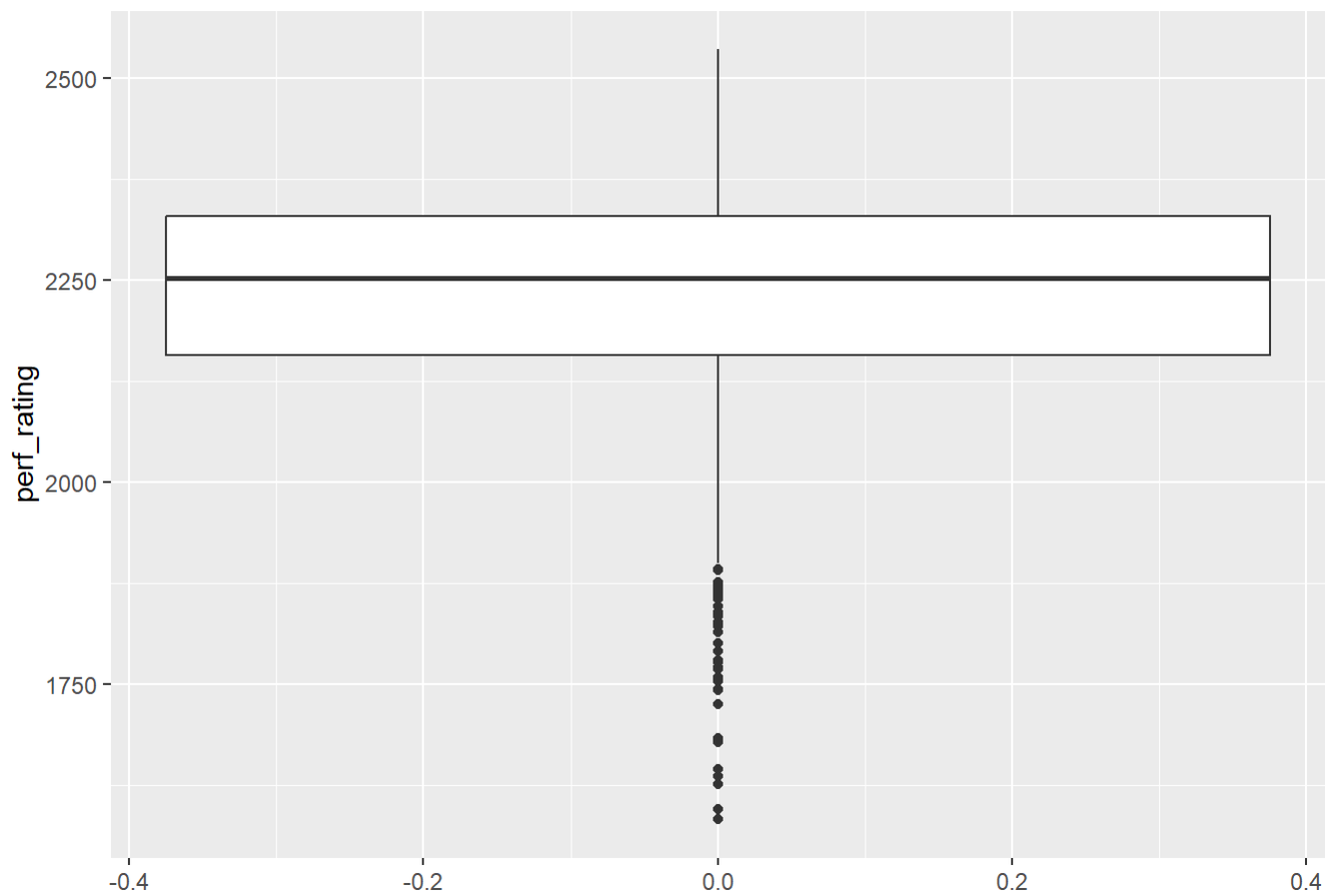
Pindi1



```
Pindi2 <- ggplot(data, aes(y = perf_rating))  
Pindi2 <- Pindi2 + geom_boxplot() + ggtitle("Repartition des rating de performance")
```

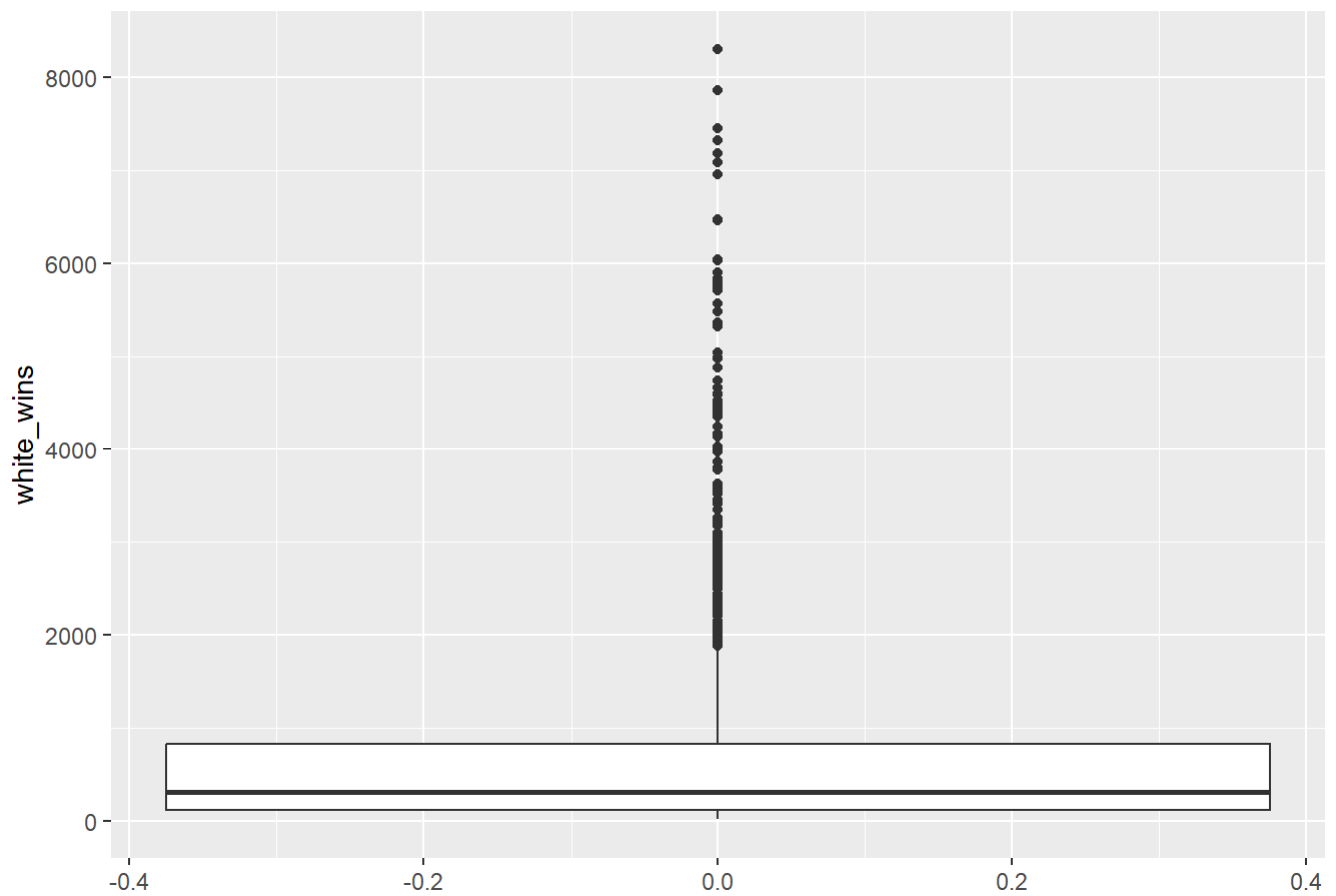
Pindi2

Repartition des rating de performance



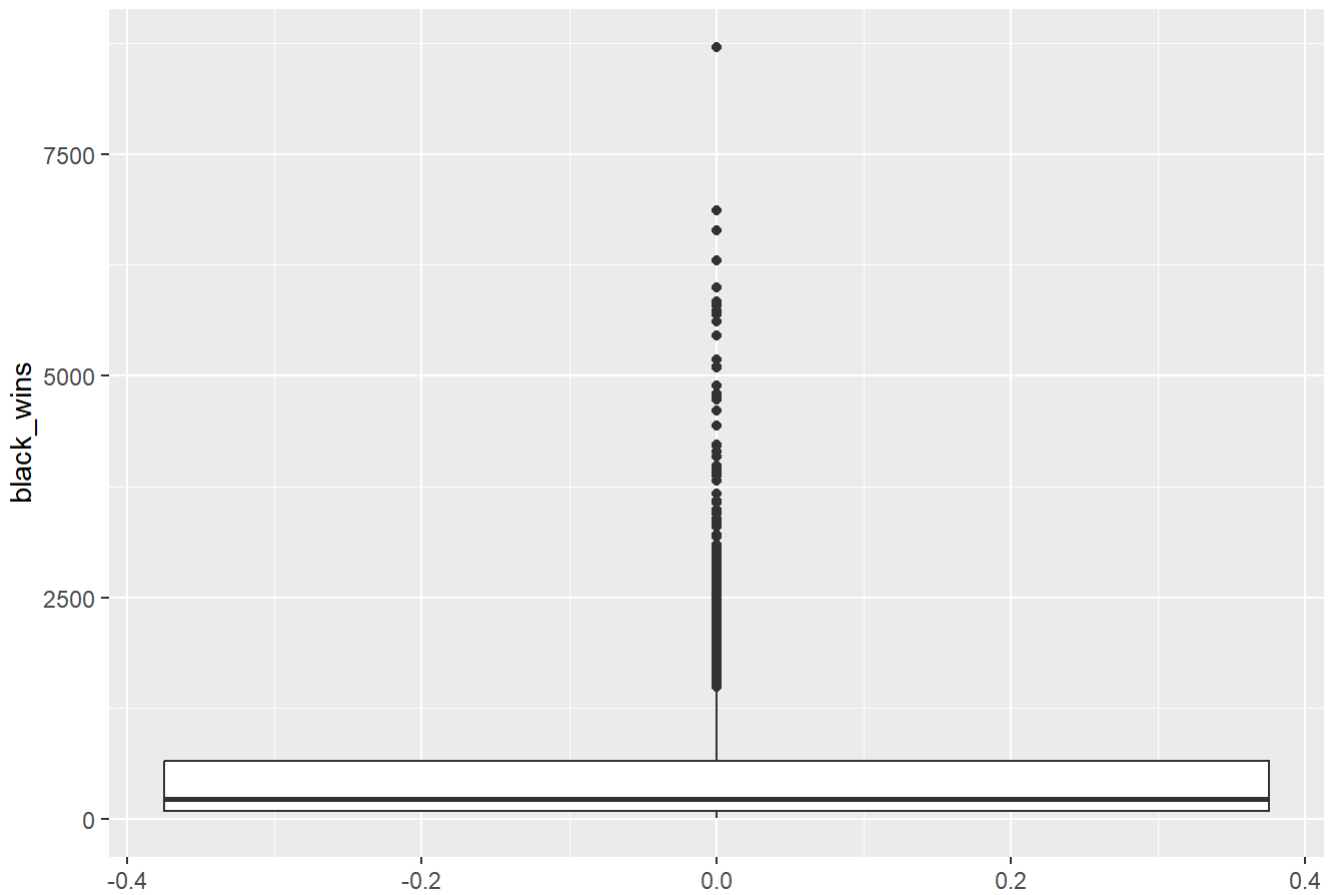
```
Pindi3 <- ggplot(data, aes(y = white_wins))  
Pindi3 <- Pindi3 + geom_boxplot() + ggtitle("Nombre de jeux gagnés par pièces blanches")  
  
Pindi3
```

Nombre de jeux gagnés par pièces blanches



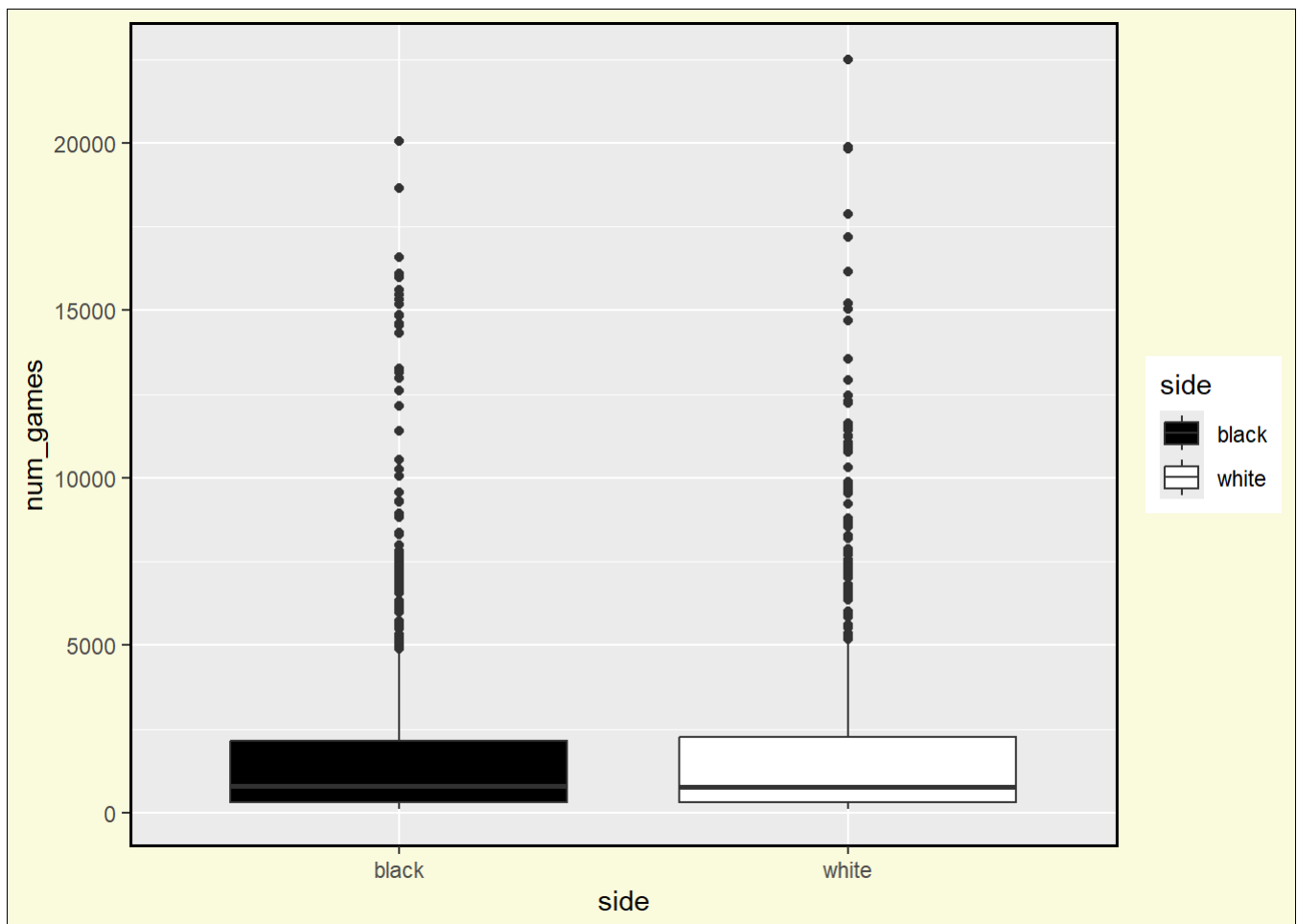
```
Pindi3 <- ggplot(data, aes(y = black_wins))  
Pindi3 <- Pindi3 + geom_boxplot() + ggtitle("Nombre de jeux gagnés par pièces noires")  
  
Pindi3
```

Nombre de jeux gagnés par pièces noires



- Graphiques combinant plusieurs variables (par exemple, avec les options facet ou facet grid de ggplot2).

```
Pcroisse1 <- data %>% ggplot(aes(x=side, y=num_games, group=side)) +  
  geom_boxplot(aes(fill=side)) + scale_fill_manual(values = c("black" = "black", "white" = "white")) +  
  theme(plot.background = element_rect(fill = "#fdfcdc", color = "black"), panel.border = element_rect(color = "black", fill = NA, size = 1)) # Changement de couleur de fond et ajouts des bordures  
Pcroisse1
```

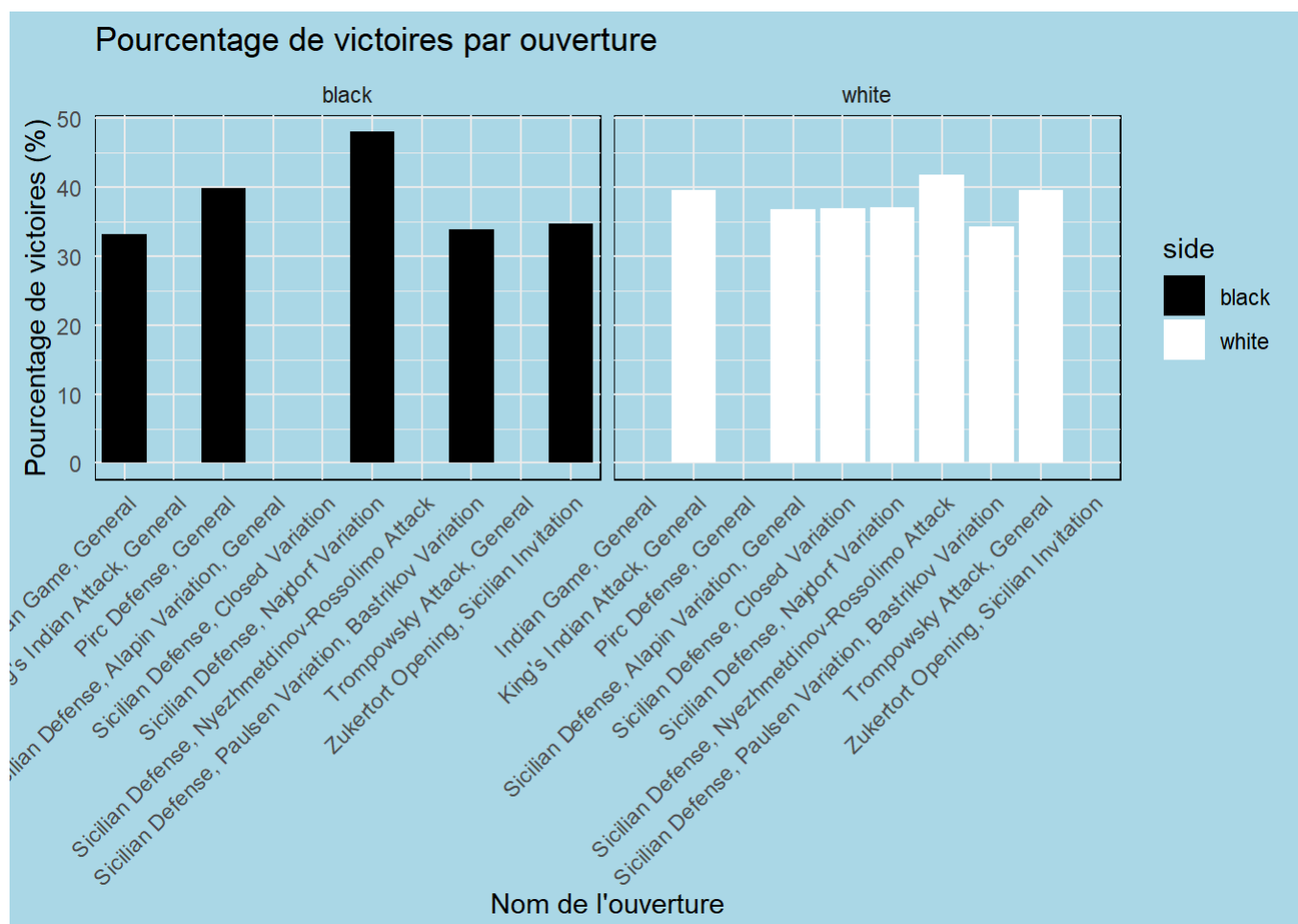
```
# Créer un graphique combiné avec facettes
ggplot(data, aes(x = num_games, y = perc_player_win, color = side)) +
  geom_point() + # Points pour représenter les données
  geom_smooth(method = "lm") + # Ajout d'une tendance linéaire
  facet_grid(. ~ side) + # Facettes par couleur (side)
  labs(
    title = "Pourcentage de victoires en fonction du nombre de parties",
    x = "Nombre de parties jouées",
    y = "Pourcentage de victoires (%)"
  ) +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Pourcentage de victoires en fonction du nombre de parties



```
# Créer un graphique combiné avec facettes
ggplot(filtered_data, aes(x = opening_name, y = perc_white_win, fill = side)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ side) + # Facetter selon la couleur jouée
  labs(
    title = "Pourcentage de victoires par ouverture",
    x = "Nom de l'ouverture",
    y = "Pourcentage de victoires (%)"
  ) +
  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "lightblue", color = NA),
    panel.background = element_rect(fill = "lightblue"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  scale_fill_manual(values = c("black" = "black", "white" = "white"))
```



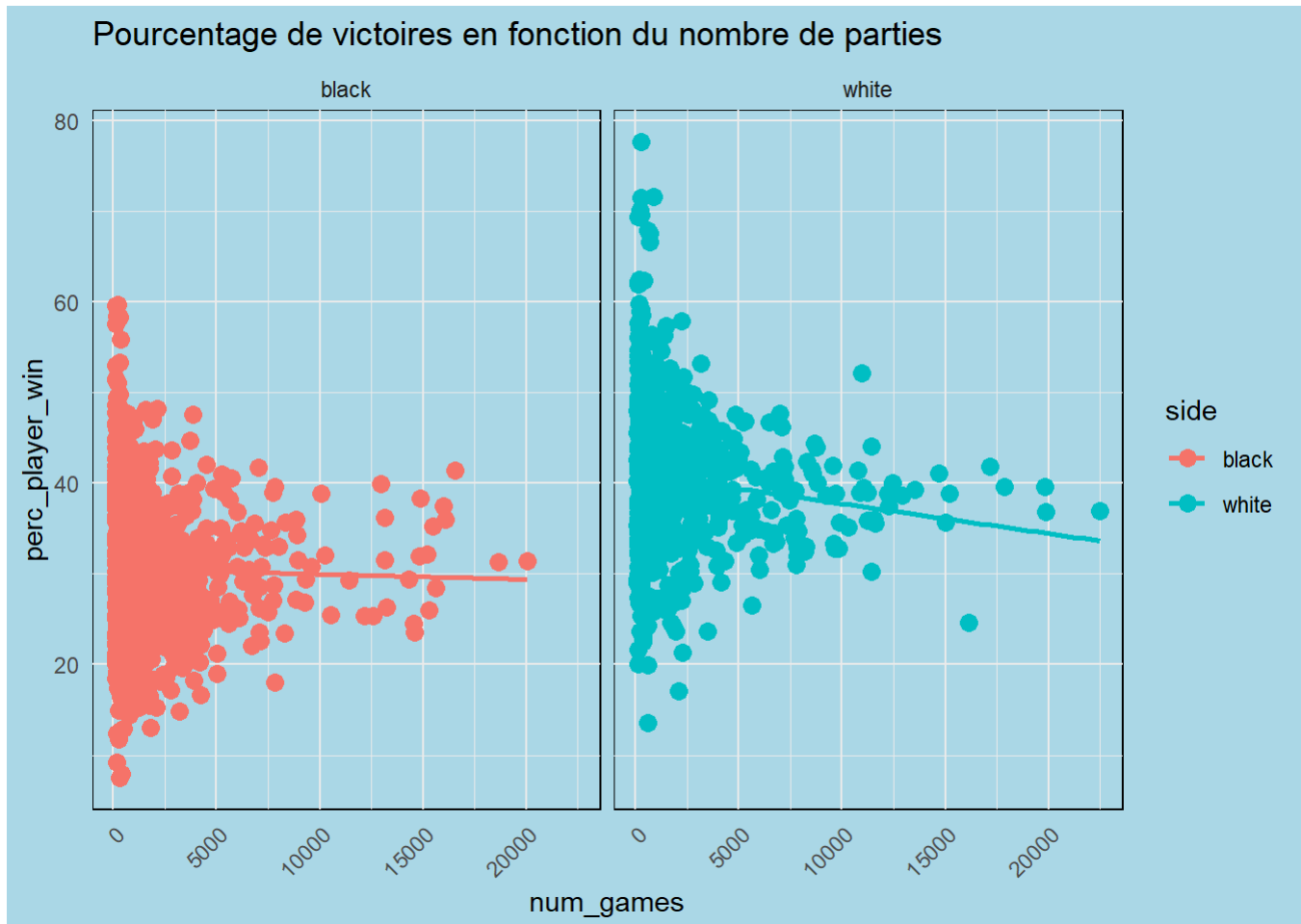
4. Programmation avancée

- Une fonction ou une boucle pour automatiser une tâche.
- Un élément de programmation avancé, tel que :
 - Une fonction complexe (par exemple une fonction à plusieurs paramètres ou une fonction permettant de générer des graphiques).
 - Une carte interactive.
 - Une interface construite avec RShiny.

```
# Appeler la fonction pour générer un graphique facetté
generate_facet_plot(
  data = data,
  x_var = "num_games",
  y_var = "perc_player_win",
  facet_var = "side",
  color_var = "side",
  title = "Pourcentage de victoires en fonction du nombre de parties",
  fill_color = "lightblue"
)
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



5. Outils recommandés

- Manipulations de données avec la librairie dplyr.
- Visualisations réalisées avec ggplot2.

6. Structure et rendu Le rendu devra être organisé sous la forme d'un projet R contenant :

- Des fichiers de code R (exemples : un fichier pour la gestion des données et un autre pour les fonctions).
- Un fichier Rmarkdown et le rapport HTML correspondant ou une application RShiny.
- Le rapport RMarkdown devra intégrer les fichiers R via la commande source().