HW1

2024-09-18

Table of Contents

[Столбчатые диаграммы 1](#_Toc178368850)

[1 1](#_Toc178368851)

[3 3](#_Toc178368852)

[Разное 4](#_Toc178368853)

[1 4](#_Toc178368854)

[2 6](#_Toc178368855)

[Боксплоты 6](#_Toc178368856)

[Фасеты 8](#_Toc178368857)

[1 8](#_Toc178368858)

[2 9](#_Toc178368859)

[3 9](#_Toc178368860)

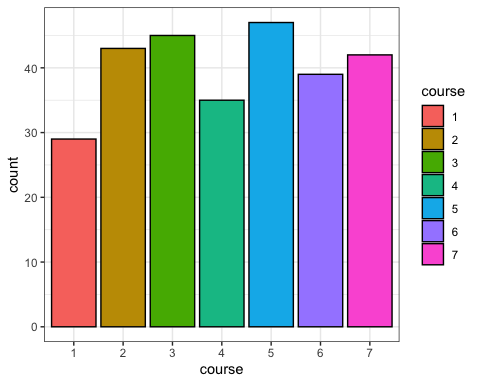
library(tidyverse)

hogwarts <- read.csv("hogwarts.csv")

## Столбчатые диаграммы

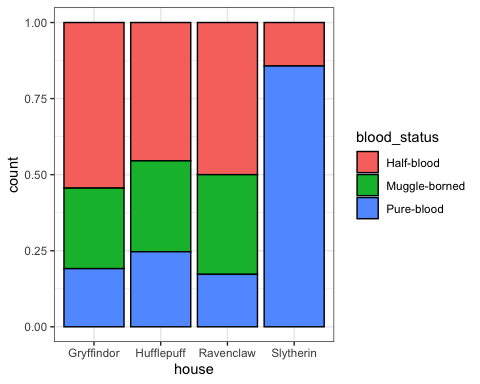
### 1

hogwarts$course <- as.factor(hogwarts$course)  
ggplot(hogwarts)+  
 geom\_bar(aes(x = course,   
 fill = course),   
 colour = "black")+  
theme\_bw()



### 2

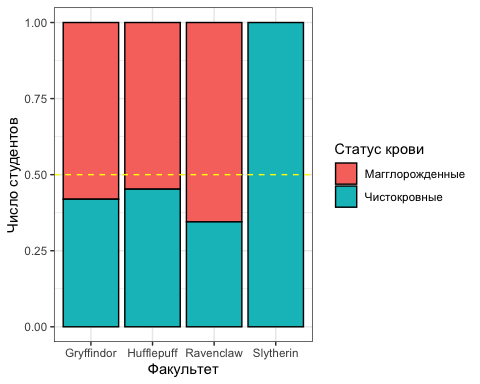
ggplot(hogwarts)+  
 geom\_bar(aes(x = house,   
 fill = blood\_status),   
 position = "fill",   
 colour = "black")+  
 theme\_bw()



На трех факультетах - Гриффиндор, Хаффлпаф и Рэйвенкло - соотношение чистокровных, полукровных и магглорожденных студентов примерно одинаковое (превалируют полукровные). На факультете Слизерин магглорожденные студенты отсутствуют, большинство - чистокровные.

### 3

hogwarts %>%  
 filter(blood\_status %in% c("Pure-blood", "Muggle-borned")) %>%  
ggplot()+  
 geom\_bar(aes(x = house,   
 fill = blood\_status),   
 position = "fill",   
 colour = "black")+  
 geom\_hline(yintercept = 0.5, linetype = "dashed", color = "yellow") +  
 scale\_x\_discrete(name = "Факультет")+  
 scale\_y\_continuous(name = "Число студентов")+  
 labs(fill = "Статус крови")+  
 scale\_fill\_discrete(labels = c("Pure-blood" = "Чистокровные", "Muggle-borned" = "Магглорожденные"))+  
 theme\_bw()



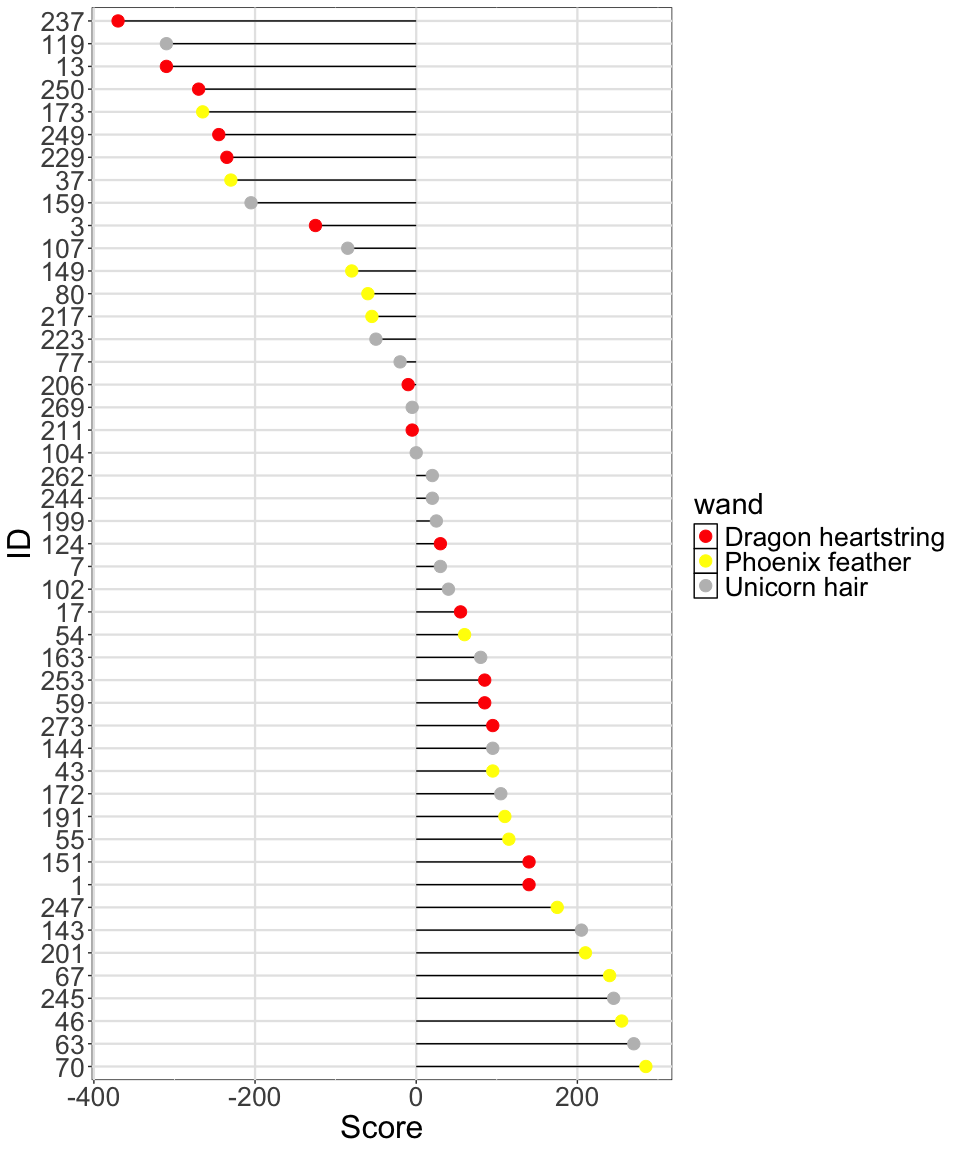
## Разное

### 1

theme\_custom <- theme(  
 title = element\_text(size = 26),   
 axis.text = element\_text(size = 20),   
 legend.text = element\_text(size = 20),   
 axis.title = element\_text(size = 24),  
 legend.title = element\_text(size = 22),  
 panel.background = element\_rect(fill = "white", colour = "black"),  
 panel.grid.major = element\_line(colour = "grey90", size = 0.8)  
 )

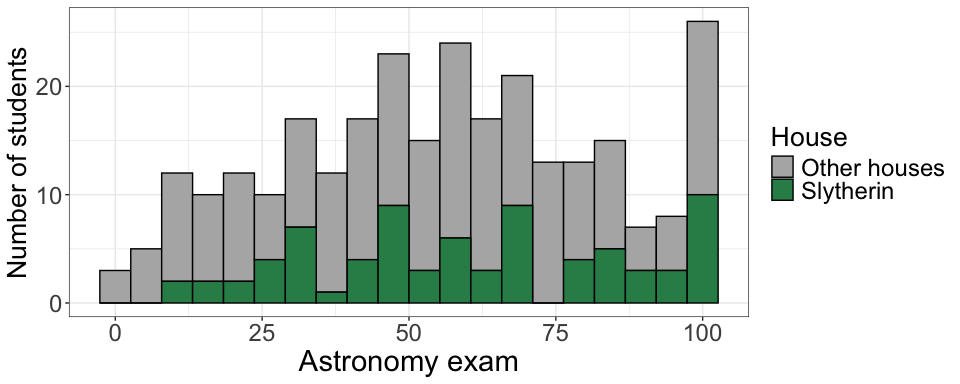
## Warning: The `size` argument of `element\_line()` is deprecated as of ggplot2 3.4.0.  
## ℹ Please use the `linewidth` argument instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

hogwarts %>%  
 filter(course == 5) %>%  
 mutate(id = as.factor(id)) %>%  
 ggplot()+  
 geom\_segment(aes(x = 0,   
 xend = result,   
 y = fct\_reorder(id, result, .desc = TRUE),   
 yend = fct\_reorder(id, result, .desc = TRUE)))+  
 geom\_point(aes(x = result, y = id,   
 colour = wand),  
 size = 4)+  
 scale\_color\_manual(values = c("Dragon heartstring" = "red",   
 "Phoenix feather" = "yellow",   
 "Unicorn hair" = "grey"))+  
 scale\_x\_continuous(name = "Score")+  
 scale\_y\_discrete(name = "ID")+  
 theme\_custom



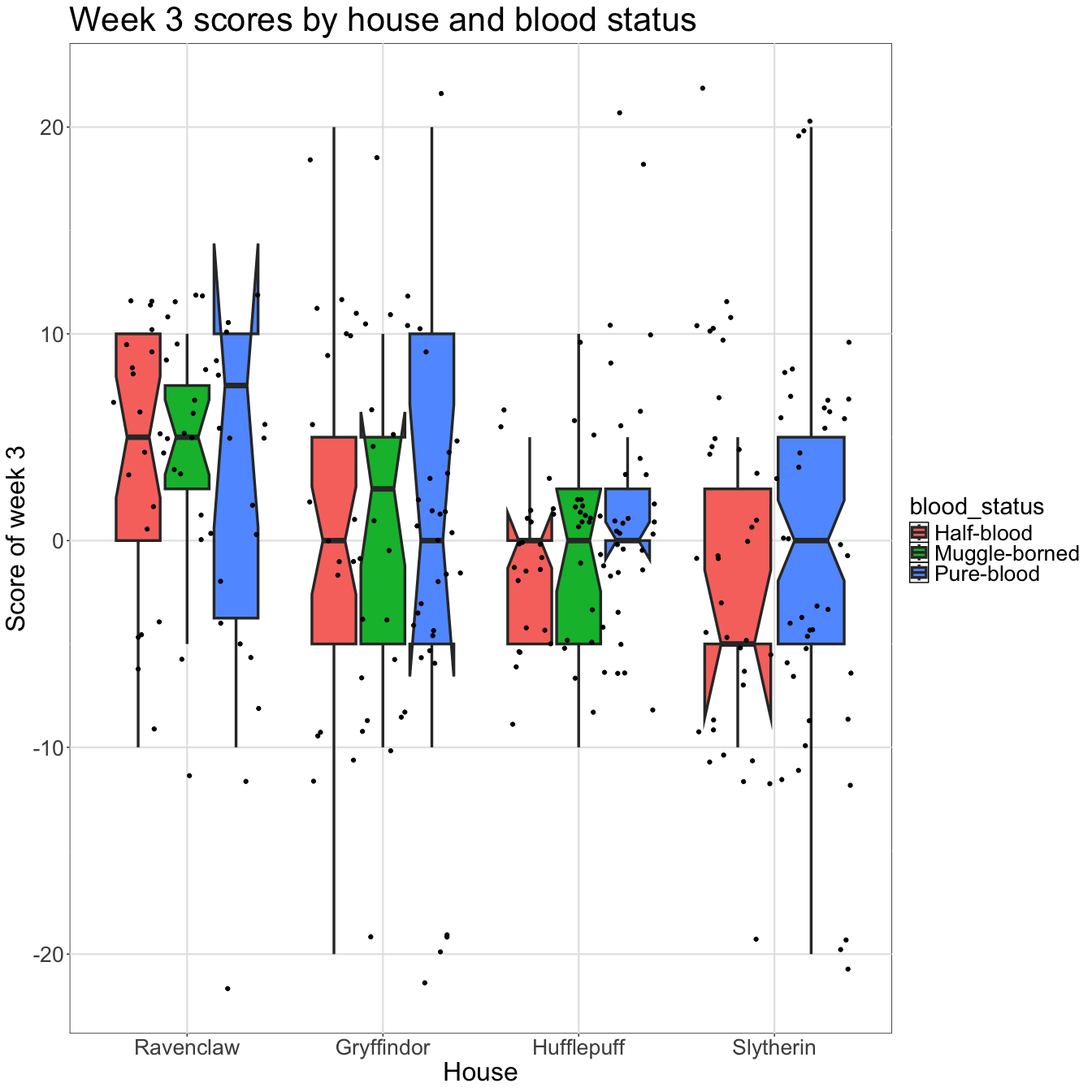
### 2

hogwarts %>%  
 ggplot()+  
 geom\_histogram(aes(x = Astronomy\_exam,  
 fill = house == "Slytherin"),  
 colour = "black",  
 bins = 20)+  
   
 labs(x = "Astronomy exam",  
 y = "Number of students",  
 fill = "House") +  
   
 scale\_fill\_manual(values = c("FALSE" = "grey70",   
 "TRUE" = "seagreen"),  
 labels = c("FALSE" = "Other houses",   
 "TRUE" = "Slytherin"))+  
 theme\_bw()+  
theme(  
 axis.text = element\_text(size = 18),  
 legend.text = element\_text(size = 18),  
 axis.title.x = element\_text(size = 22),  
 axis.title.y = element\_text(size = 20),  
 legend.title = element\_text(size = 20))



## Боксплоты

theme\_custom <- theme(  
 title = element\_text(size = 26),   
 axis.text = element\_text(size = 20),   
 legend.text = element\_text(size = 20),   
 axis.title = element\_text(size = 24),  
 legend.title = element\_text(size = 22),  
 panel.background = element\_rect(fill = "white", colour = "black"),  
 panel.grid.major = element\_line(colour = "grey90", size = 0.8)  
 )  
  
hogwarts %>%  
 select(id, `house`, `blood\_status`, `week.3`) %>%  
ggplot(aes(x = fct\_reorder(house, week.3, .desc = TRUE),   
 y = week.3))+  
 geom\_boxplot(aes(fill = blood\_status),  
 notch = TRUE,  
 outliers = FALSE,  
 width = 0.75,   
 size = 1.2)+  
 geom\_jitter()+  
   
 labs(title = "Week 3 scores by house and blood status",   
 x = "House",   
 y = "Score of week 3")+  
   
theme\_custom



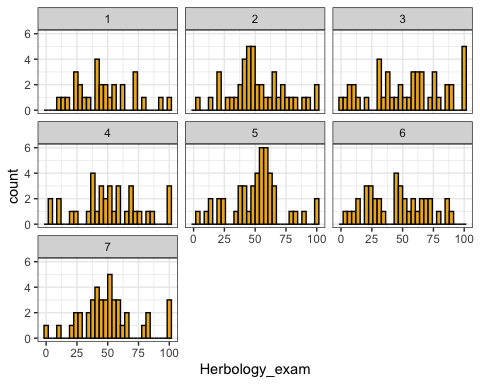
## Фасеты

### 1

На мой взгляд, способы фасетирования себя проявляют по разному. В случае фасетирования по строкам нагляднее сравнивается распределение групп по количественной переменной (по баллам например). В случае фасетирования по столбцам - не сжимается высота столбцов и лучше визуально оценивается частота данных, распределение по переменной тоже видно, но хуже. Так что в общем случае предпочтительнее будет распределение по столбцам. Для визуаализации violin-plotов тоже лучше использовать фасетирование вертикальное (по столбцам), лучше видно и ширину вайолинов и можно визуально сравнить расположение толстой части (медианы). В общем случае лучше использовать фасетирование по столбцам. Фасетирование по строкам можно использовать, когда нужно сравнить распределение групп по количественной переменной и при этом частота данных либо не важна, либо особо не отличается в группах.

### 2

hogwarts %>%  
ggplot(aes(x = Herbology\_exam))+  
 geom\_histogram(bins = 30, fill = "goldenrod2", colour = "black")+  
 #facet\_grid(.~course)+  
 facet\_wrap(vars(course))+  
 #geom\_violin(aes(y = course))+  
   
theme\_bw()



### 3

hogwarts %>%  
 ggplot()+  
 geom\_density(aes(x = Defence.against.the.dark.arts\_exam),   
 fill = "lightpink2",  
 alpha = 0.5  
 )+  
 geom\_density(aes(x = Herbology\_exam),  
 fill = "dodgerblue",  
 alpha = 0.5  
 )+  
 facet\_grid(.~sex)+  
   
theme(strip.text = element\_text(size = 20))+  
theme\_custom

