Вasic\_visualization\_DZ1

Saldaeva V

2024-09-21

## Data reading

hogwarts <- read\_csv("C:/Users/Алексей/Desktop/mag/statistics/biostatistics/визуализация данных/data/hogwarts\_2024.csv")  
hogwarts |> head()

## # A tibble: 6 × 60  
## id house course sex wandCore bloodStatus result Defence against the …¹  
## <dbl> <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl>  
## 1 1 Ravencl… 4 fema… unicorn… half-blood 94 73  
## 2 2 Hufflep… 5 male phoenix… half-blood 33 38  
## 3 3 Ravencl… 4 fema… dragon … half-blood 137 52  
## 4 4 Hufflep… 2 male phoenix… half-blood 27 50  
## 5 5 Hufflep… 2 fema… phoenix… half-blood 67 47  
## 6 6 Gryffin… 6 male phoenix… muggle-born 126 44  
## # ℹ abbreviated name: ¹​`Defence against the dark arts exam`  
## # ℹ 52 more variables: `Flying exam` <dbl>, `Astronomy exam` <dbl>,  
## # `Herbology exam` <dbl>, `Divinations exam` <dbl>, `Charms exam` <dbl>,  
## # `History of magic exam` <dbl>, `Arithmancy exam` <dbl>,  
## # `Care of magical creatures exam` <dbl>, `Muggle studies exam` <dbl>,  
## # `Study of ancient runes exam` <dbl>, `Transfiguration exam` <dbl>,  
## # `Potions exam` <dbl>, week\_1 <dbl>, week\_2 <dbl>, week\_3 <dbl>, …

### Checking dataset structure

hogwarts |> glimpse()

## Rows: 560  
## Columns: 60  
## $ id <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11…  
## $ house <chr> "Ravenclaw", "Hufflepuff", "Raven…  
## $ course <dbl> 4, 5, 4, 2, 2, 6, 7, 5, 2, 3, 7, …  
## $ sex <chr> "female", "male", "female", "male…  
## $ wandCore <chr> "unicorn hair", "phoenix feather"…  
## $ bloodStatus <chr> "half-blood", "half-blood", "half…  
## $ result <dbl> 94, 33, 137, 27, 67, 126, 63, 7, …  
## $ `Defence against the dark arts exam` <dbl> 73, 38, 52, 50, 47, 44, 51, 47, 2…  
## $ `Flying exam` <dbl> 33, 36, 73, 42, 41, 52, 34, 34, 2…  
## $ `Astronomy exam` <dbl> 57, 45, 66, 49, 57, 59, 58, 37, 5…  
## $ `Herbology exam` <dbl> 73, 50, 62, 39, 38, 46, 59, 23, 2…  
## $ `Divinations exam` <dbl> 66, 54, 72, 42, 47, 49, 42, 38, 1…  
## $ `Charms exam` <dbl> 60, 70, 77, 46, 35, 55, 86, 20, 4…  
## $ `History of magic exam` <dbl> 52, 36, 60, 45, 50, 40, 55, 21, 2…  
## $ `Arithmancy exam` <dbl> 61, 36, 58, 32, 76, 50, 41, 31, 2…  
## $ `Care of magical creatures exam` <dbl> 44, 41, 70, 36, 46, 73, 29, 36, 4…  
## $ `Muggle studies exam` <dbl> 64, 34, 52, 59, 50, 54, 36, 31, 4…  
## $ `Study of ancient runes exam` <dbl> 50, 35, 59, 39, 48, 56, 47, 41, 3…  
## $ `Transfiguration exam` <dbl> 74, 70, 70, 15, 32, 86, 100, 31, …  
## $ `Potions exam` <dbl> 67, 38, 22, 64, 56, 60, 62, 55, 1…  
## $ week\_1 <dbl> 0, -5, 0, -1, 1, 5, 1, -20, 3, -2…  
## $ week\_2 <dbl> -10, 1, 0, 5, 20, 10, -5, 10, 1, …  
## $ week\_3 <dbl> 0, -1, 1, -5, 10, -5, 3, -5, -3, …  
## $ week\_4 <dbl> 10, 1, -1, 10, -10, 10, 0, -10, -…  
## $ week\_5 <dbl> 3, -5, 3, 0, -1, 20, 5, 5, -3, 5,…  
## $ week\_6 <dbl> -20, 20, 0, 0, 0, 0, 0, 5, 0, -1,…  
## $ week\_7 <dbl> 10, 10, 1, -3, -20, 1, 10, 3, -5,…  
## $ week\_8 <dbl> 5, 5, 1, -5, 5, 5, 0, 1, 0, 20, -…  
## $ week\_9 <dbl> 1, 1, 3, -1, 0, 3, -20, -20, -10,…  
## $ week\_10 <dbl> 20, -10, 1, 5, -1, 0, 5, -5, 5, 3…  
## $ week\_11 <dbl> 5, -10, 20, 0, 0, 0, 5, 10, 5, 5,…  
## $ week\_12 <dbl> 5, -5, 1, -20, -10, -5, 0, 5, 1, …  
## $ week\_13 <dbl> -20, -5, 10, 0, 0, 1, -1, 10, -20…  
## $ week\_14 <dbl> 0, 5, 3, 10, -10, 20, 0, -20, -20…  
## $ week\_15 <dbl> 1, 20, 1, 0, -20, 10, 1, 3, -20, …  
## $ week\_16 <dbl> 20, 5, 5, 5, 0, 3, 10, -1, 5, 5, …  
## $ week\_17 <dbl> 3, 0, 10, 5, 5, -5, -1, 10, -10, …  
## $ week\_18 <dbl> 10, 5, 5, 5, 10, -20, 0, 10, 3, 5…  
## $ week\_19 <dbl> -10, 0, -5, -1, 0, -1, 0, 20, 0, …  
## $ week\_20 <dbl> 10, -10, 5, 10, 0, -1, -1, 10, 0,…  
## $ week\_21 <dbl> 0, 5, 5, 3, 5, 0, 0, -5, -5, 5, 5…  
## $ week\_22 <dbl> 20, -5, 5, 0, 20, 5, -1, 0, 0, 20…  
## $ week\_23 <dbl> 5, 1, -3, 20, -5, 20, 0, 1, 1, 5,…  
## $ week\_24 <dbl> 10, -20, -20, 0, 10, 5, 5, -3, -5…  
## $ week\_25 <dbl> 0, -20, 1, 3, 5, 1, -5, 0, -20, 2…  
## $ week\_26 <dbl> 10, 10, 5, -1, 0, 5, 5, -3, 0, 20…  
## $ week\_27 <dbl> 5, 5, -3, 0, 20, 5, 0, -5, 10, 3,…  
## $ week\_28 <dbl> -3, 20, 20, 1, 10, 5, 1, 10, 0, 1…  
## $ week\_29 <dbl> -20, -5, 5, 5, -10, 1, 0, -3, 0, …  
## $ week\_30 <dbl> 5, 1, -5, 5, -5, -1, -20, 20, 1, …  
## $ week\_31 <dbl> 5, 5, 20, -5, -10, -3, 0, -10, 20…  
## $ week\_32 <dbl> -5, 1, 20, -1, -10, 5, 10, 1, 0, …  
## $ week\_33 <dbl> 0, 10, 3, 3, 0, 0, -1, 0, -20, 3,…  
## $ week\_34 <dbl> 0, -1, 0, 0, 10, 3, 20, -5, 10, 3…  
## $ week\_35 <dbl> 5, -5, 3, -10, 3, -5, 0, 0, 0, 0,…  
## $ week\_36 <dbl> 1, 5, 1, -20, 5, 20, -1, -3, 1, 3…  
## $ week\_37 <dbl> 0, 0, 10, -1, 10, 3, 3, 0, 20, 1,…  
## $ week\_38 <dbl> 10, -1, 0, -5, 5, 5, 20, -5, -3, …  
## $ week\_39 <dbl> 3, 5, 1, 10, 20, 0, 5, 1, -5, 0, …  
## $ week\_40 <dbl> 0, 0, 5, 1, 5, 1, 10, -5, -20, 3,…

# Changing some variables type to factors  
hogwarts <- hogwarts |> mutate(  
 across(c(house, course, sex, wandCore, bloodStatus), ~ as.factor(.x))  
)

### NA checking

sum(is.na(hogwarts))

## [1] 0

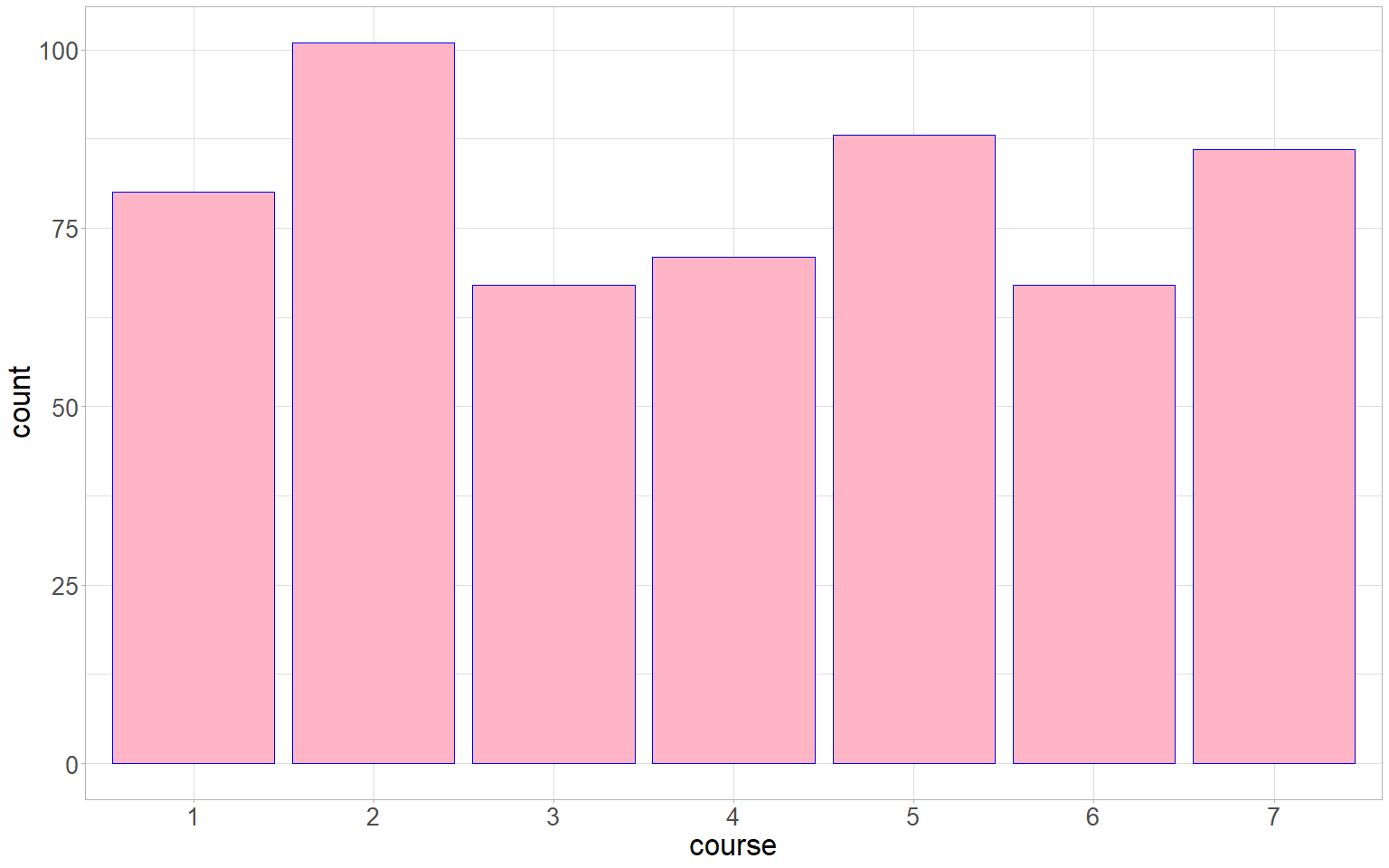
### Summary output

hogwarts |> summary()

## id house course sex wandCore   
## Min. : 1.0 Gryffindor:126 1: 80 female:333 dragon heartstring:196   
## 1st Qu.:140.8 Hufflepuff:179 2:101 male :227 phoenix feather :181   
## Median :280.5 Ravenclaw :122 3: 67 unicorn hair :183   
## Mean :280.5 Slytherin :133 4: 71   
## 3rd Qu.:420.2 5: 88   
## Max. :560.0 6: 67   
## 7: 86   
## bloodStatus result Defence against the dark arts exam  
## half-blood :391 Min. :-292.00 Min. : 0   
## muggle-born: 60 1st Qu.: 7.00 1st Qu.:39   
## pure-blood :109 Median : 70.50 Median :49   
## Mean : 59.71 Mean :48   
## 3rd Qu.: 128.25 3rd Qu.:58   
## Max. : 260.00 Max. :89   
##   
## Flying exam Astronomy exam Herbology exam Divinations exam  
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.00   
## 1st Qu.:36.00 1st Qu.:37.00 1st Qu.:39.00 1st Qu.:38.00   
## Median :48.00 Median :49.00 Median :49.00 Median :49.00   
## Mean :47.37 Mean :47.99 Mean :47.75 Mean :48.44   
## 3rd Qu.:60.00 3rd Qu.:60.00 3rd Qu.:58.00 3rd Qu.:59.00   
## Max. :85.00 Max. :87.00 Max. :86.00 Max. :89.00   
##   
## Charms exam History of magic exam Arithmancy exam  
## Min. : 0.00 Min. : 0.00 Min. : 0.00   
## 1st Qu.:39.00 1st Qu.:37.00 1st Qu.:38.00   
## Median :49.00 Median :48.00 Median :50.00   
## Mean :48.36 Mean :47.28 Mean :48.38   
## 3rd Qu.:59.00 3rd Qu.:58.00 3rd Qu.:60.00   
## Max. :98.00 Max. :85.00 Max. :91.00   
##   
## Care of magical creatures exam Muggle studies exam Study of ancient runes exam  
## Min. : 0.00 Min. : 0.00 Min. : 0.00   
## 1st Qu.:38.00 1st Qu.:38.00 1st Qu.:38.00   
## Median :49.00 Median :50.00 Median :48.00   
## Mean :48.11 Mean :48.64 Mean :47.44   
## 3rd Qu.:60.00 3rd Qu.:61.00 3rd Qu.:58.00   
## Max. :95.00 Max. :94.00 Max. :89.00   
##   
## Transfiguration exam Potions exam week\_1 week\_2   
## Min. : 0.00 Min. : 0.00 Min. :-20.000 Min. :-20.000   
## 1st Qu.: 34.00 1st Qu.: 21.00 1st Qu.: -3.000 1st Qu.: -3.000   
## Median : 49.00 Median : 47.00 Median : 1.000 Median : 1.000   
## Mean : 48.24 Mean : 46.62 Mean : 1.334 Mean : 1.161   
## 3rd Qu.: 62.25 3rd Qu.: 68.00 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. :100.00 Max. :100.00 Max. : 50.000 Max. : 20.000   
##   
## week\_3 week\_4 week\_5 week\_6   
## Min. :-20.000 Min. :-20.00 Min. :-20.0000 Min. :-20.000   
## 1st Qu.: -1.500 1st Qu.: -1.00 1st Qu.: -3.0000 1st Qu.: -1.000   
## Median : 1.000 Median : 1.00 Median : 1.0000 Median : 1.000   
## Mean : 1.407 Mean : 1.82 Mean : 0.9196 Mean : 1.448   
## 3rd Qu.: 5.000 3rd Qu.: 5.00 3rd Qu.: 5.0000 3rd Qu.: 5.000   
## Max. : 20.000 Max. : 20.00 Max. : 20.0000 Max. : 20.000   
##   
## week\_7 week\_8 week\_9 week\_10   
## Min. :-20.000 Min. :-20.0 Min. :-50.00 Min. :-20.000   
## 1st Qu.: -3.000 1st Qu.: -1.0 1st Qu.: -1.00 1st Qu.: -1.000   
## Median : 1.000 Median : 1.0 Median : 1.00 Median : 1.000   
## Mean : 1.529 Mean : 1.6 Mean : 1.63 Mean : 1.457   
## 3rd Qu.: 5.000 3rd Qu.: 5.0 3rd Qu.: 5.00 3rd Qu.: 5.000   
## Max. : 20.000 Max. : 20.0 Max. : 20.00 Max. : 20.000   
##   
## week\_11 week\_12 week\_13 week\_14   
## Min. :-20.000 Min. :-20.000 Min. :-20.0000 Min. :-20.00   
## 1st Qu.: -1.000 1st Qu.: -1.000 1st Qu.: -3.0000 1st Qu.: -1.00   
## Median : 1.000 Median : 1.000 Median : 0.0000 Median : 1.00   
## Mean : 1.586 Mean : 1.689 Mean : 0.7393 Mean : 1.53   
## 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 5.0000 3rd Qu.: 5.00   
## Max. : 20.000 Max. : 20.000 Max. : 50.0000 Max. : 20.00   
##   
## week\_15 week\_16 week\_17 week\_18   
## Min. :-20.000 Min. :-20.000 Min. :-20.0 Min. :-20.000   
## 1st Qu.: -1.000 1st Qu.: -1.000 1st Qu.: -1.0 1st Qu.: -1.000   
## Median : 1.000 Median : 1.000 Median : 1.0 Median : 1.000   
## Mean : 1.738 Mean : 1.636 Mean : 1.8 Mean : 1.712   
## 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 5.0 3rd Qu.: 5.000   
## Max. : 20.000 Max. : 20.000 Max. : 50.0 Max. : 20.000   
##   
## week\_19 week\_20 week\_21 week\_22   
## Min. :-50.0000 Min. :-20.00 Min. :-20.000 Min. :-20.000   
## 1st Qu.: -3.0000 1st Qu.: -3.00 1st Qu.: -1.000 1st Qu.: -1.000   
## Median : 0.0000 Median : 1.00 Median : 1.000 Median : 1.000   
## Mean : 0.8071 Mean : 1.55 Mean : 1.816 Mean : 1.527   
## 3rd Qu.: 5.0000 3rd Qu.: 5.00 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. : 20.0000 Max. : 50.00 Max. : 20.000 Max. : 20.000   
##   
## week\_23 week\_24 week\_25 week\_26   
## Min. :-20.0000 Min. :-20.000 Min. :-20.000 Min. :-20.000   
## 1st Qu.: -3.0000 1st Qu.: -1.000 1st Qu.: -3.000 1st Qu.: -3.000   
## Median : 0.0000 Median : 1.000 Median : 1.000 Median : 1.000   
## Mean : 0.8036 Mean : 1.168 Mean : 1.364 Mean : 1.248   
## 3rd Qu.: 5.0000 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. : 20.0000 Max. : 20.000 Max. : 20.000 Max. : 20.000   
##   
## week\_27 week\_28 week\_29 week\_30   
## Min. :-50.0 Min. :-20.000 Min. :-20.000 Min. :-20.000   
## 1st Qu.: -1.0 1st Qu.: -1.500 1st Qu.: -1.000 1st Qu.: -1.000   
## Median : 1.0 Median : 1.000 Median : 0.000 Median : 1.000   
## Mean : 1.5 Mean : 1.923 Mean : 1.262 Mean : 1.705   
## 3rd Qu.: 5.0 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. : 20.0 Max. : 20.000 Max. : 20.000 Max. : 20.000   
##   
## week\_31 week\_32 week\_33 week\_34   
## Min. :-20.00 Min. :-20.000 Min. :-20.000 Min. :-20.000   
## 1st Qu.: -1.00 1st Qu.: -1.000 1st Qu.: -1.000 1st Qu.: -1.000   
## Median : 1.00 Median : 1.000 Median : 1.000 Median : 1.000   
## Mean : 1.68 Mean : 2.013 Mean : 1.539 Mean : 1.593   
## 3rd Qu.: 5.00 3rd Qu.: 5.000 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. : 20.00 Max. : 20.000 Max. : 20.000 Max. : 20.000   
##   
## week\_35 week\_36 week\_37 week\_38   
## Min. :-20.0 Min. :-20.000 Min. :-20.00 Min. :-20.000   
## 1st Qu.: -1.0 1st Qu.: -1.000 1st Qu.: -1.00 1st Qu.: -1.000   
## Median : 1.0 Median : 1.000 Median : 1.00 Median : 1.000   
## Mean : 1.7 Mean : 2.079 Mean : 1.32 Mean : 1.864   
## 3rd Qu.: 5.0 3rd Qu.: 5.000 3rd Qu.: 5.00 3rd Qu.: 5.000   
## Max. : 20.0 Max. : 20.000 Max. : 20.00 Max. : 20.000   
##   
## week\_39 week\_40   
## Min. :-20.000 Min. :-20.000   
## 1st Qu.: -1.000 1st Qu.: -3.000   
## Median : 1.000 Median : 0.000   
## Mean : 1.438 Mean : 1.079   
## 3rd Qu.: 5.000 3rd Qu.: 5.000   
## Max. : 20.000 Max. : 20.000   
##

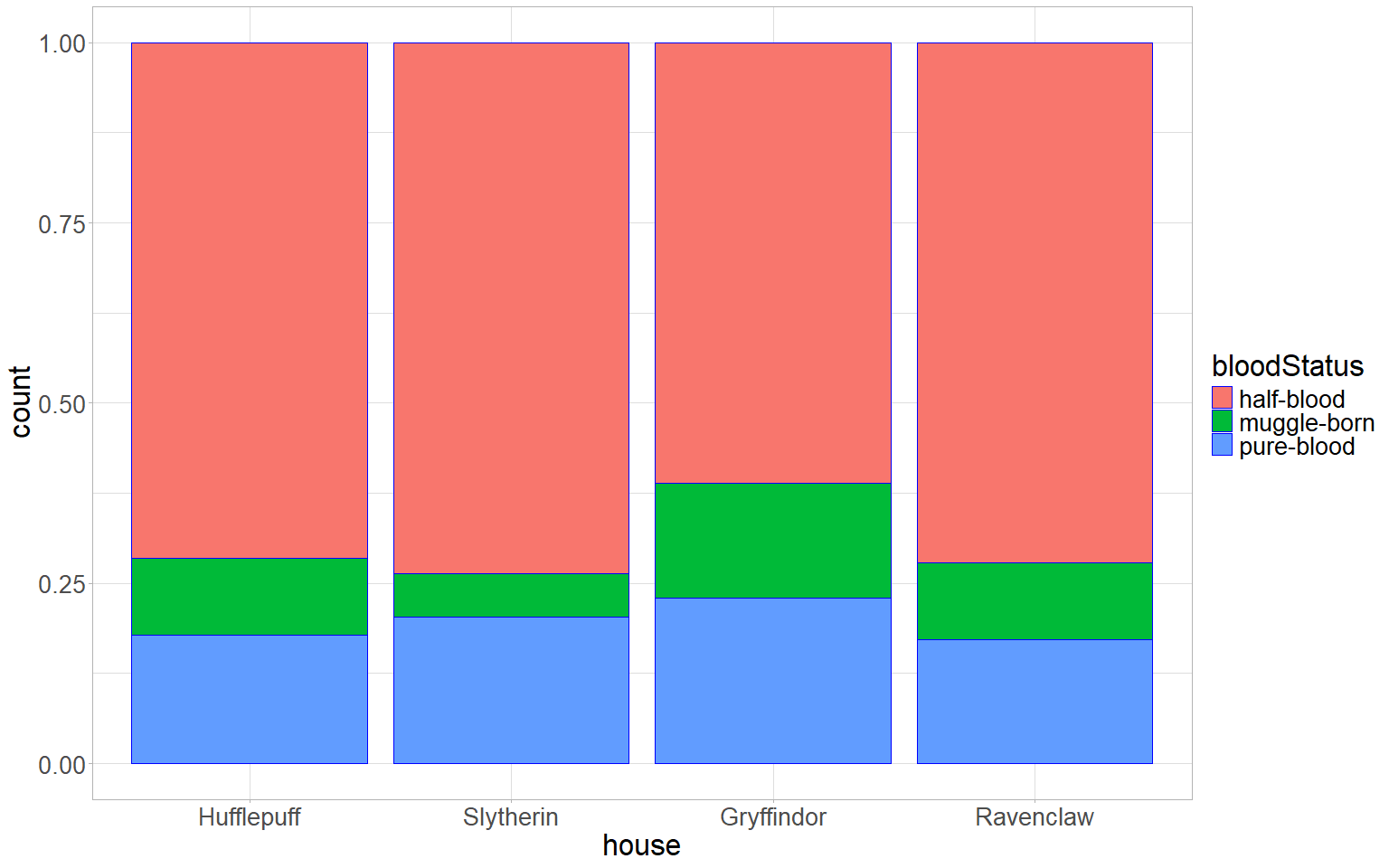
###Bar charts ## Data visualisation number student/courses

ggplot(hogwarts) +  
 geom\_bar(aes(x = course),  
 fill = "pink1",  
 colour = "blue")+  
 theme\_light()+  
 theme(  
 axis.text = element\_text(size = 20),  
 axis.title = element\_text(size = 25)  
 )



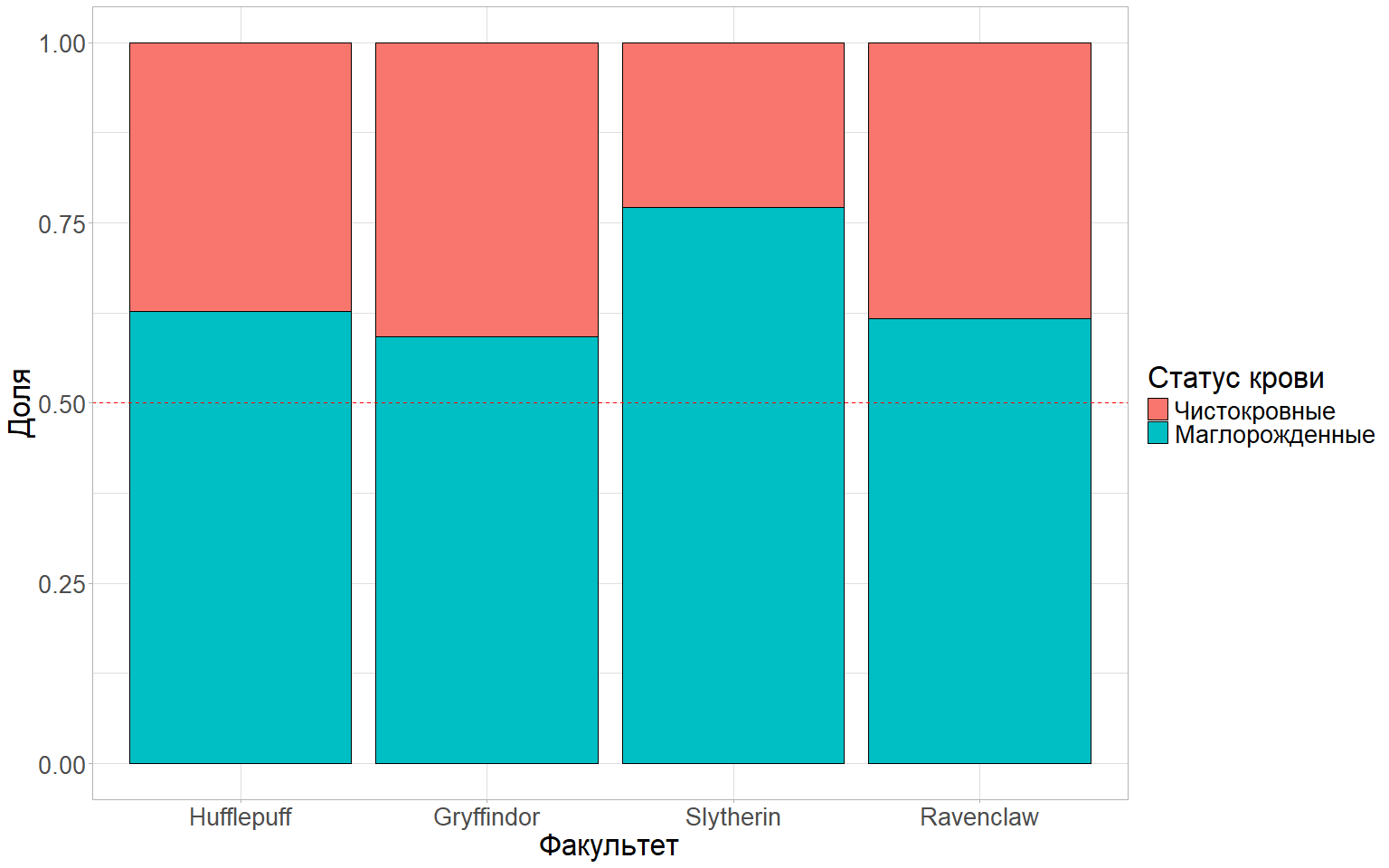
# number of students by house/bloodStatus.

ggplot(hogwarts) +  
 geom\_bar(aes(x = fct\_infreq(house),   
 fill = bloodStatus),  
 position = "fill",  
 colour = "blue")+   
 scale\_x\_discrete(name = "house")+   
 theme\_light()+  
 theme(  
 axis.text = element\_text(size = 20),  
 axis.title = element\_text(size = 25),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25)  
 )

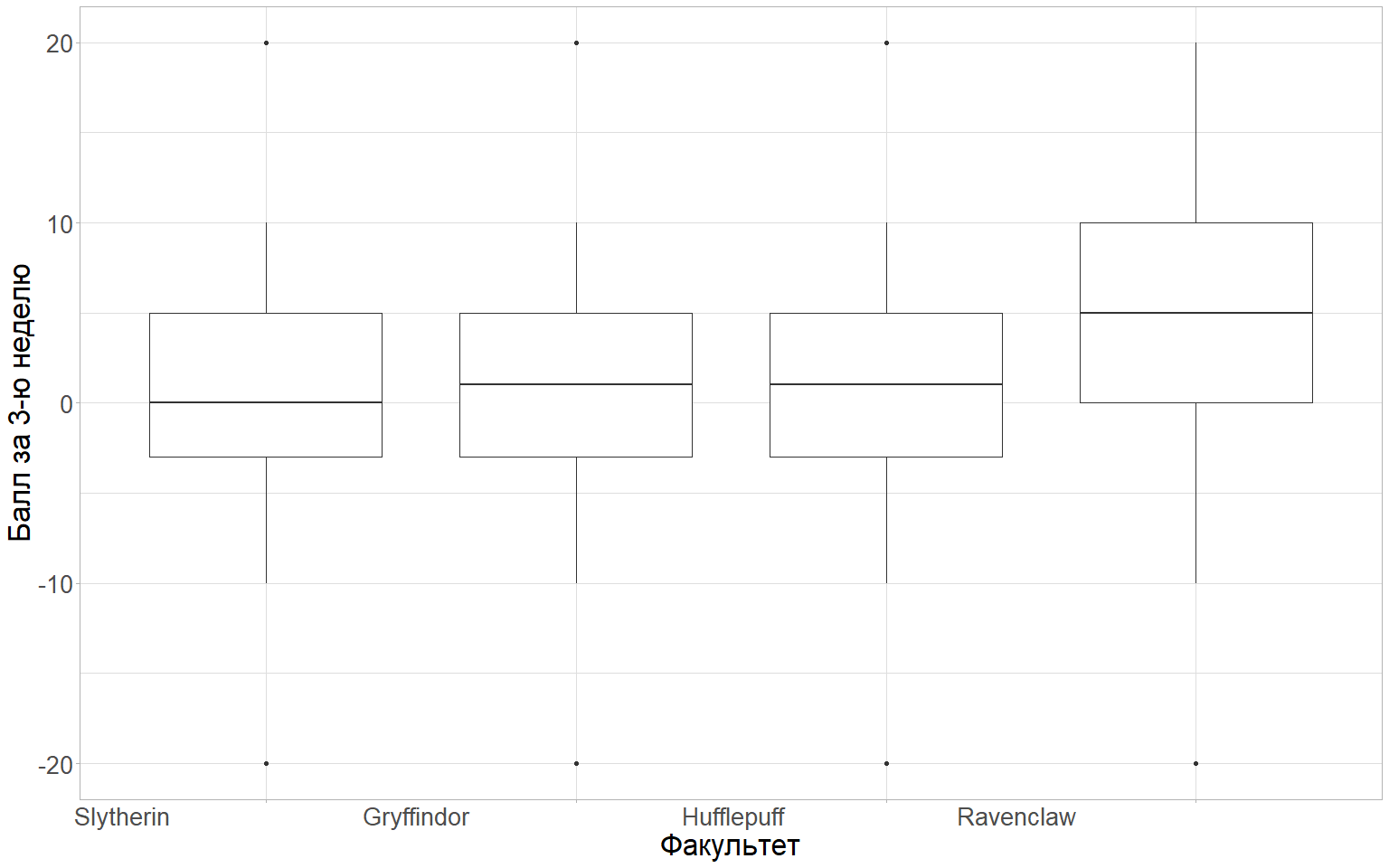


## Modification of the dataset

hogwarts\_mod <- hogwarts %>%  
 filter(bloodStatus %in% c("pure-blood", "muggle-born"))  
ggplot(hogwarts\_mod, aes(x = fct\_infreq(house), fill = bloodStatus)) +  
 geom\_bar(position = "fill",  
 colour = "black") +  
 geom\_hline(yintercept = 0.5, linetype = "dashed", color = "red") +  
 scale\_fill\_discrete(name = "Статус крови", labels = c("Чистокровные", "Маглорожденные")) +  
 xlab("Факультет") +  
 ylab("Доля") +  
 theme\_light() +  
 theme(  
 axis.text = element\_text(size = 20),  
 axis.title = element\_text(size = 25),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25)  
 )

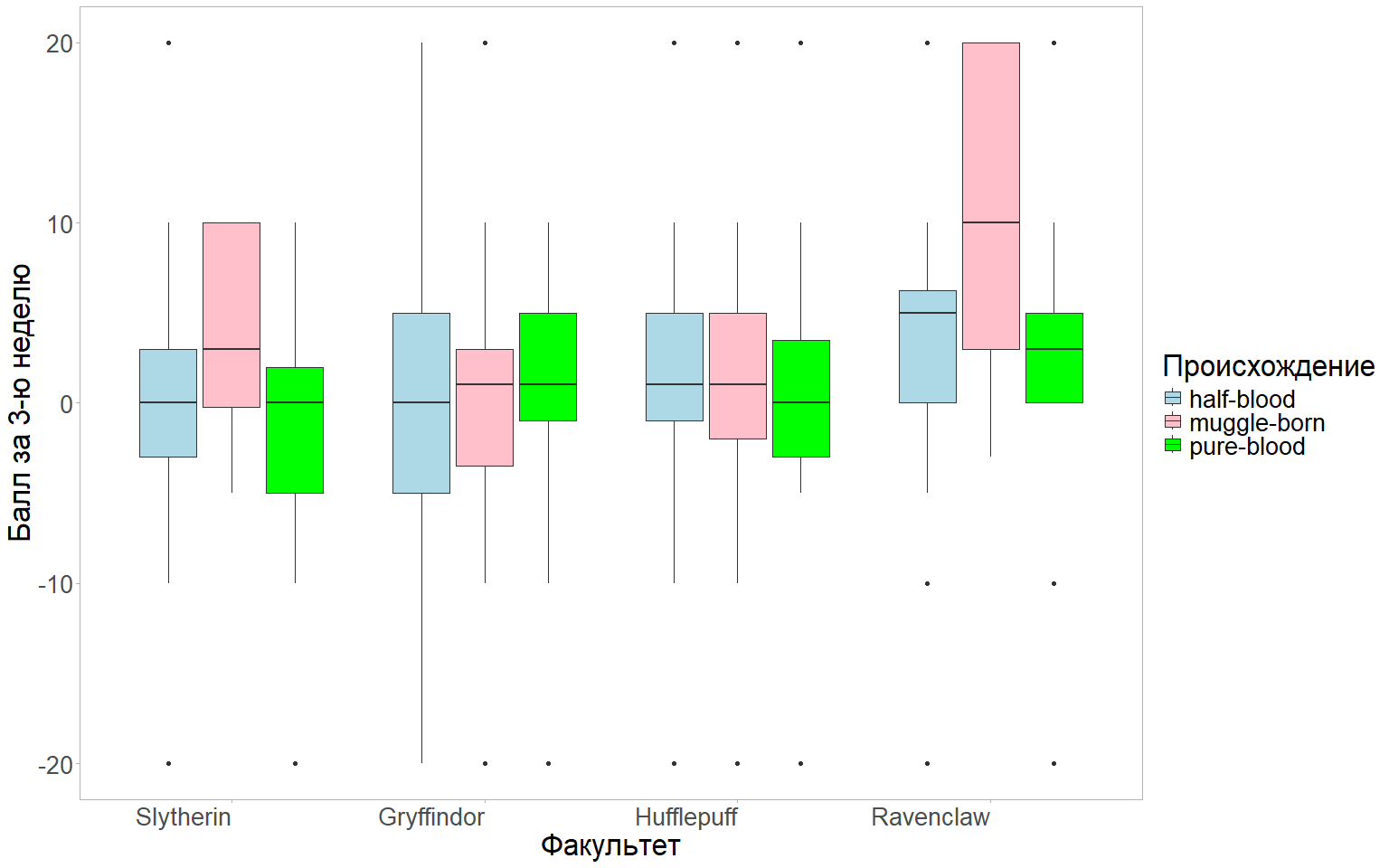
 ### Boxplots ## The distribution of points earned by students in the 3rd week of study, by faculty.

ggplot(hogwarts, aes(x = fct\_reorder(house, week\_3), y = week\_3)) +  
 geom\_boxplot() +  
 labs(x = "Факультет", y = "Балл за 3-ю неделю") +  
 theme\_light() +  
 theme(  
 axis.text.x = element\_text(angle = 0, hjust = 2),  
 axis.title = element\_text(size = 25),  
 axis.text = element\_text(size = 20)  
 )

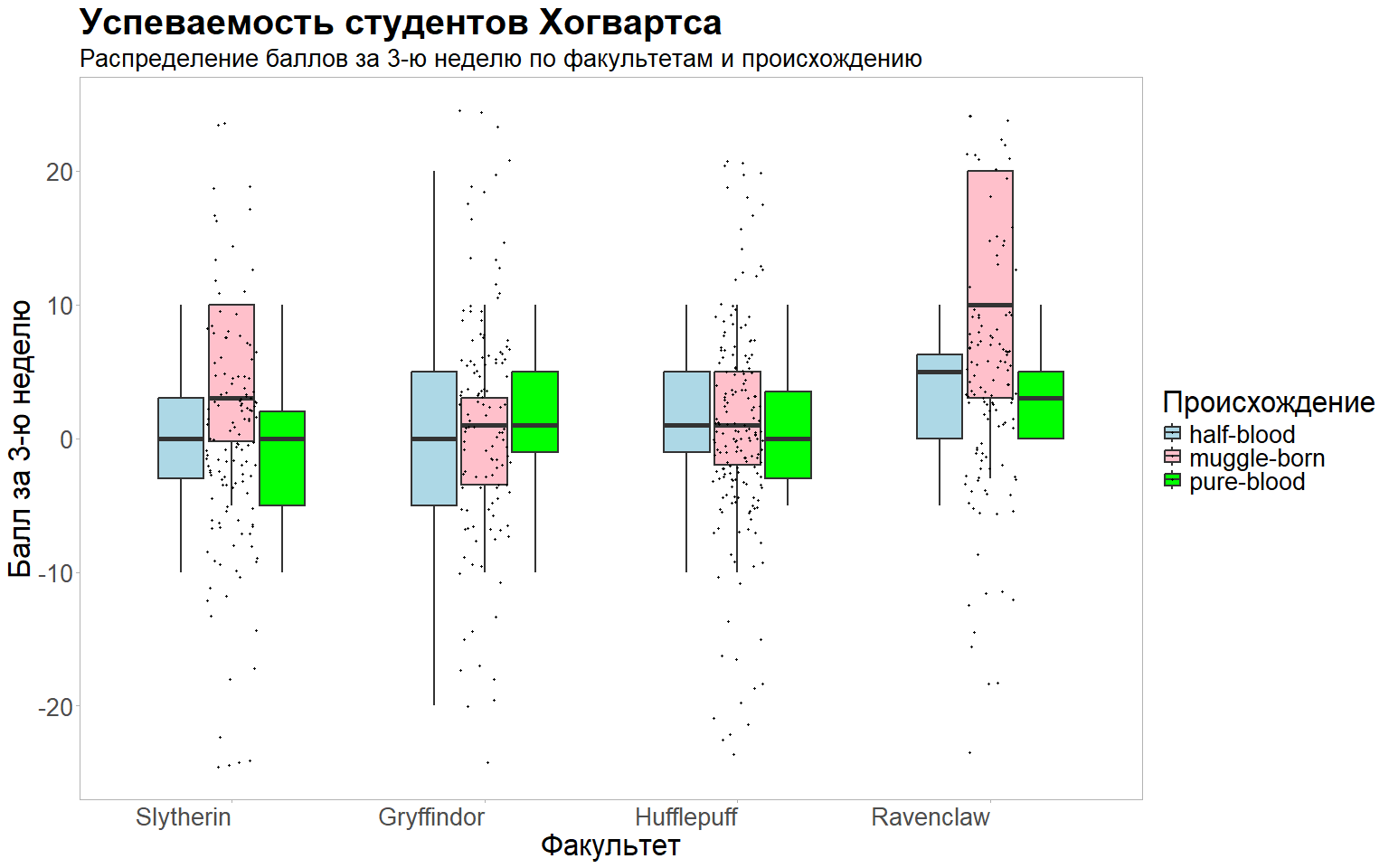


## Added a display in different colors for the origin of students (bloodStatus). I added a notch to the boxplot.

ggplot(hogwarts, aes(x = fct\_reorder(house, week\_3), y = week\_3, fill = bloodStatus)) +  
 geom\_boxplot(notch = FALSE) +  
 labs(x = "Факультет", y = "Балл за 3-ю неделю", fill = "Происхождение") +  
 theme\_light() +  
 theme(  
 axis.text.x = element\_text(angle = 90, hjust = 1),  
 axis.title = element\_text(size = 14),  
 axis.text = element\_text(size = 14)  
 ) +  
 scale\_fill\_manual(values = c("half-blood" = "lightblue", "muggle-born" = "pink", "pure-blood" = "green")) +  
 theme(plot.background = element\_rect(fill = "white", color = NA),  
 panel.background = element\_rect(fill = "white", color = NA),  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank()) +  
 theme(plot.height = 14, plot.width = 14) +  
theme(  
 axis.text.x = element\_text(angle = 0, hjust = 1),  
 axis.title = element\_text(size = 25),  
 axis.text = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25)  
 )

 ## I added a jitter raft to the chart. Removed the outlier display from the boxplot.

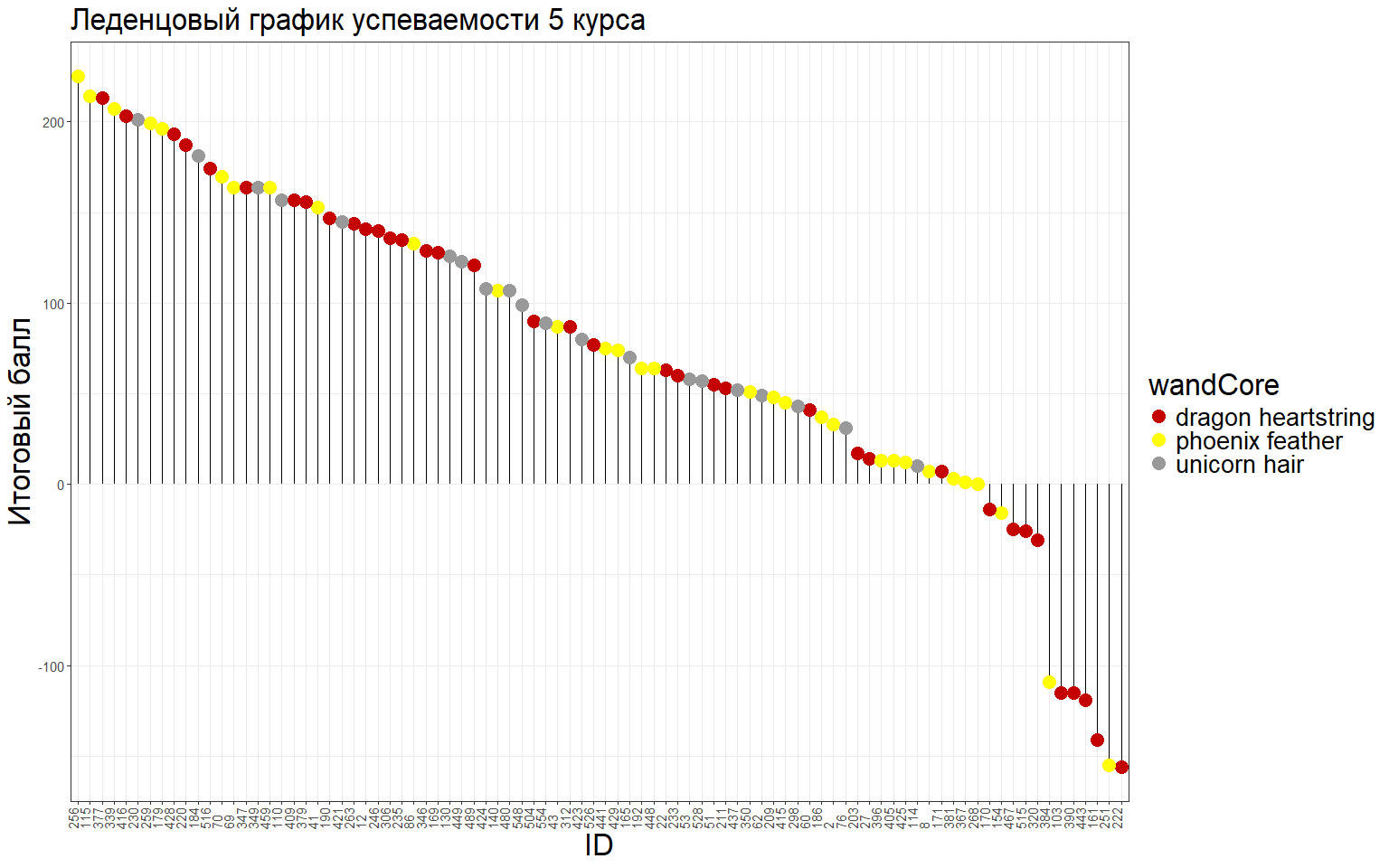
ggplot(hogwarts, aes(x = fct\_reorder(house, week\_3), y = week\_3, fill = bloodStatus)) +  
 geom\_boxplot(notch = FALSE, outlier.shape = NA, linewidth = 1.0, width = 0.6) +  
 geom\_jitter(shape = 20, size = 1.5, position = position\_jitter(width = 0.1, height = 5)) +  
 labs(x = "Факультет", y = "Балл за 3-ю неделю", fill = "Происхождение", title = "Успеваемость студентов Хогвартса", subtitle = "Распределение баллов за 3-ю неделю по факультетам и происхождению") +  
 theme\_light() +  
 theme(  
 axis.text.x = element\_text(angle = 0, hjust = 1, size = 20),  
 axis.title = element\_text(size = 25),  
 axis.text = element\_text(size = 20),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25),  
 plot.title = element\_text(size = 30, face = "bold"),  
 plot.subtitle = element\_text(size = 20)  
 ) +  
 scale\_fill\_manual(values = c("half-blood" = "lightblue", "muggle-born" = "pink", "pure-blood" = "green")) +  
 theme(plot.background = element\_rect(fill = "white", color = NA),  
 panel.background = element\_rect(fill = "white", color = NA),  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank()) +  
 theme(plot.height = 14, plot.width = 14)



### Various

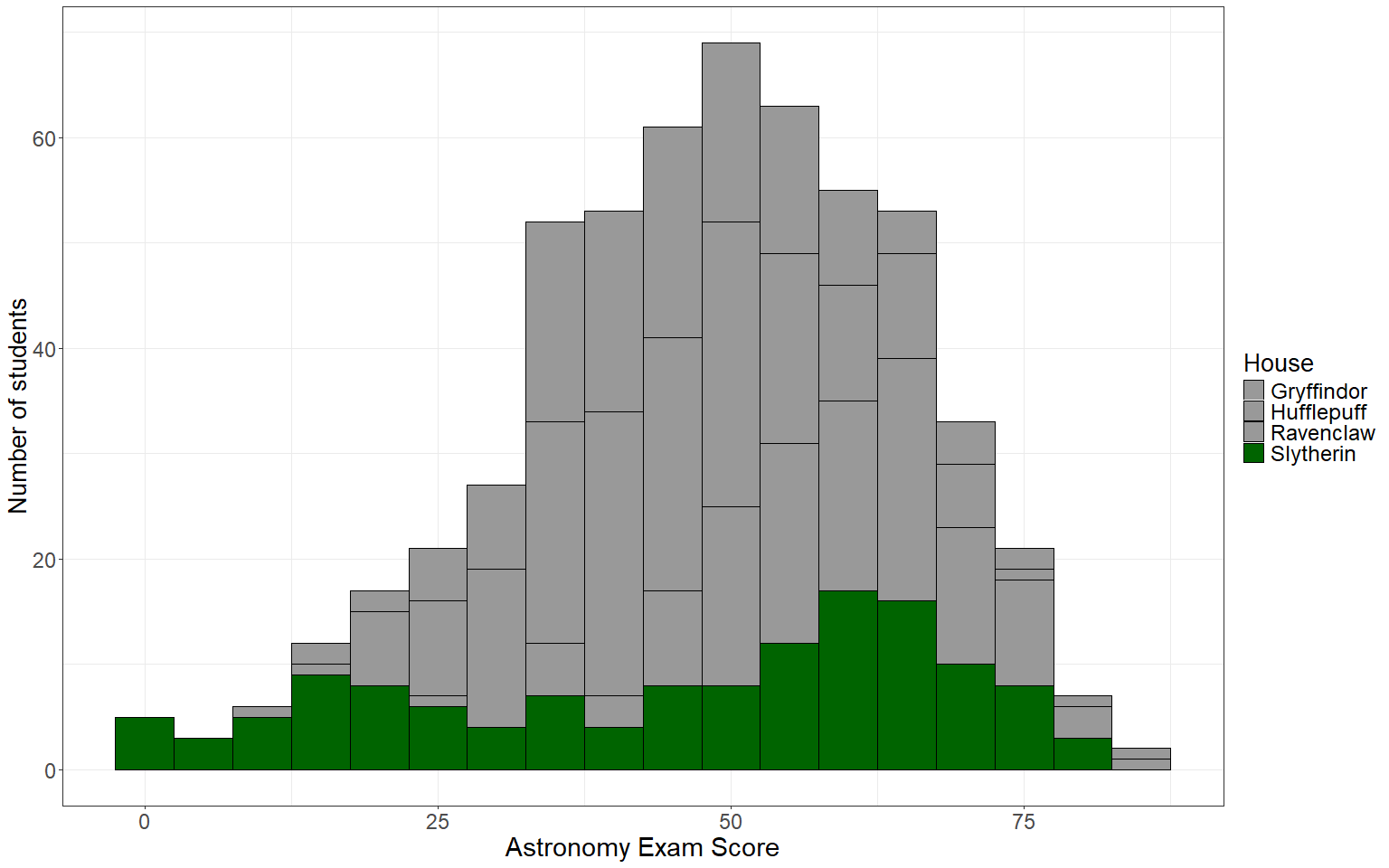
## 1

hogwarts %>%   
 filter(course == 5) %>%  
 mutate(id = as.factor(id)) %>%  
 ggplot() +  
 geom\_segment(aes(x = fct\_reorder(id, result, .desc = TRUE), xend = id, y = 0, yend = result)) +  
 geom\_point(aes(x = id, y = result, color = wandCore), size = 5) +  
 scale\_color\_manual(values = c("dragon heartstring" = "#C50000", "phoenix feather" = "#FFFF00", "unicorn hair" = "#999999")) +  
 labs(x = "id", y = "Итоговый балл", title = "Леденцовый график успеваемости 5 курса") +  
 scale\_x\_discrete(name = "ID") +  
 theme\_bw() +  
 theme(  
 axis.text.x = element\_text(angle = 90, vjust = 0,5, hjust = 0),  
 axis.text = element\_text(size = 11),  
 axis.title = element\_text(size = 25),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25),  
 plot.title = element\_text(size = 25)  
 )



## 2

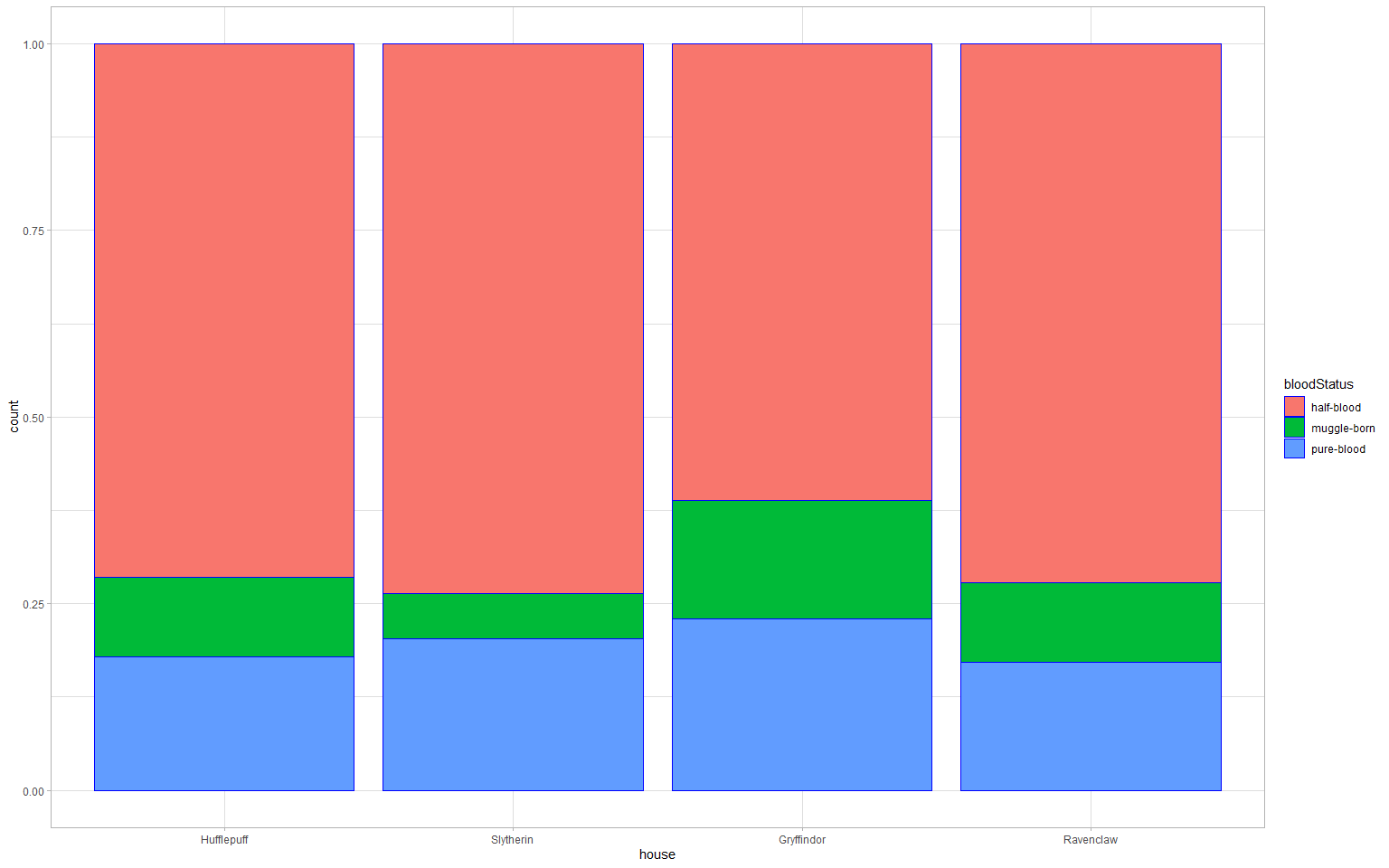
ggplot(hogwarts, aes(x = `Astronomy exam`, fill = factor(house))) +  
 geom\_histogram(binwidth = 5, color = "black") +  
 scale\_fill\_manual(values = c("Gryffindor" = "#999999", "Hufflepuff" = "#999999", "Ravenclaw" = "#999999", "Slytherin" = "#006400")) +  
 labs(x = "Astronomy Exam Score", y = "Number of students", fill = "House") +  
 theme\_bw() +  
 theme(  
 axis.text = element\_text(size = 18),  
 axis.title.x = element\_text(size = 22),  
 axis.title.y = element\_text(size = 20),  
 legend.text = element\_text(size = 18),  
 legend.title = element\_text(size = 20)  
 )



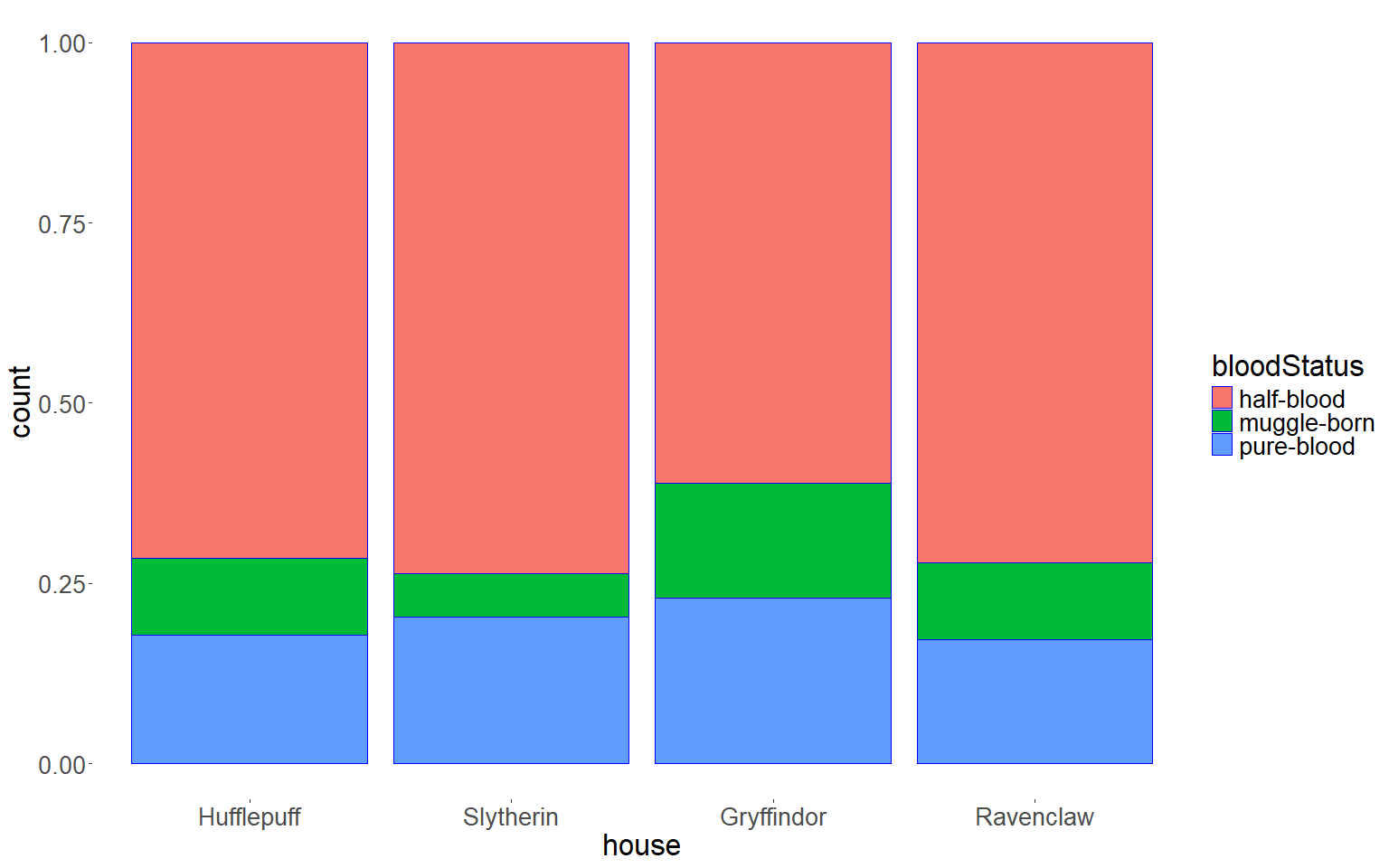
### 3

theme\_custom <- function() {  
 theme(  
 panel.background = element\_rect(fill = "white"),  
 axis.text = element\_text(size = 20),  
 axis.title = element\_text(size = 25),  
 legend.text = element\_text(size = 20),  
 legend.title = element\_text(size = 25)  
 )  
}

ggplot(hogwarts, aes(x = fct\_infreq(house), fill = bloodStatus)) +  
 geom\_bar(position = "fill", colour = "blue") +  
 scale\_x\_discrete(name = "house") +  
 theme\_light()



# Применим новую кастомную тему  
ggplot(hogwarts, aes(x = fct\_infreq(house), fill = bloodStatus)) +  
 geom\_bar(position = "fill", colour = "blue") +  
 scale\_x\_discrete(name = "house") +  
 theme\_custom()



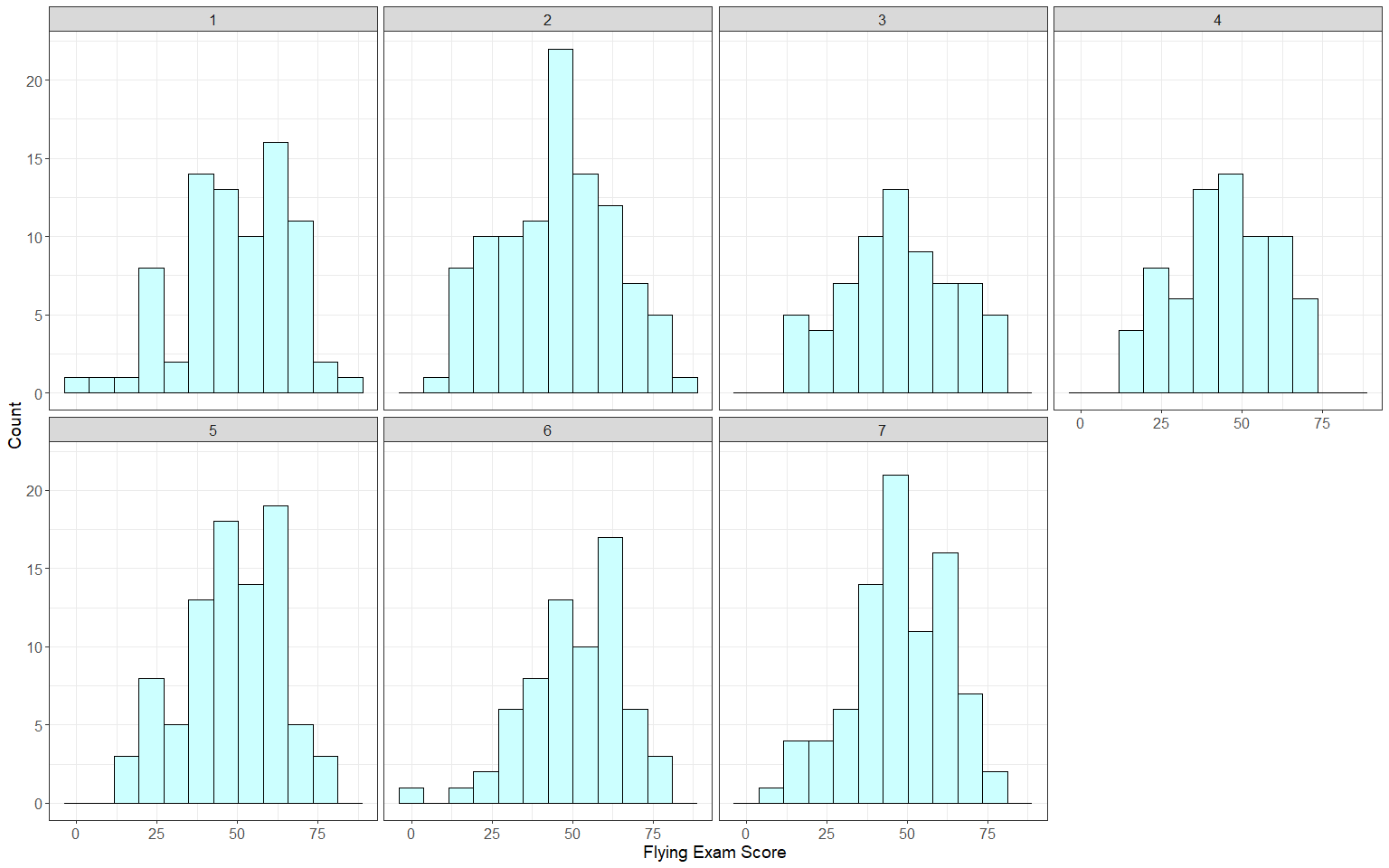
### Faceting

## 1

Напишите, какой, по вашему мнению, способ фасетирования (по строкам или по столбцам) лучше использовать для визуализации гистограммы. Фасетирование по столбцам для гистограмм удобнее для восприятия, т.к читается привычным способом слева направо, поэтому визуальная информация считывается быстрее. А какой для визуализации violin-plot? На мой взгляд, фасетирование по столбцам для violin-plot более удобно, по причине удобства восприятия информации справа налево и хорошо воспринимается демонстрация сравнений между несколькими категориями,а также можно использовать цветовые палитры(как и в гистограммах).

## 2

ggplot(hogwarts, aes(x = `Flying exam`)) +  
 geom\_histogram(bins = 12, color = "black", fill = "#CCFFFF") +  
 facet\_wrap(~ course, ncol = 4) +  
 labs(x = "Flying Exam Score", y = "Count") +  
 theme\_bw() +  
 theme(  
 axis.text = element\_text(size = 12),  
 axis.title = element\_text(size = 14),  
 strip.text = element\_text(size = 12)  
 )



hogwarts\_clean <- hogwarts %>%   
 filter(!is.na(`Defence against the dark arts exam`), !is.na(`Herbology exam`))  
  
# Создаем график  
ggplot(hogwarts\_clean) +  
 geom\_density(aes(x = `Defence against the dark arts exam`, fill = "Defence against the dark arts exam"), alpha = 0.5) +  
 geom\_density(aes(x = `Herbology exam`, fill = "Herbology exam"), alpha = 0.5) +  
 scale\_fill\_manual(values = c("Defence against the dark arts exam" = "blue", "Herbology exam" = "green")) +  
 facet\_wrap(~ sex) + # фасетирование по полу  
 labs(title = "Распределение плотности вероятности по экзаменам",  
 x = "Оценка",  
 y = "Плотность вероятности",  
 fill = "Экзамены") +  
 theme\_custom()

