RUDN university

5100

Программирование в среде R

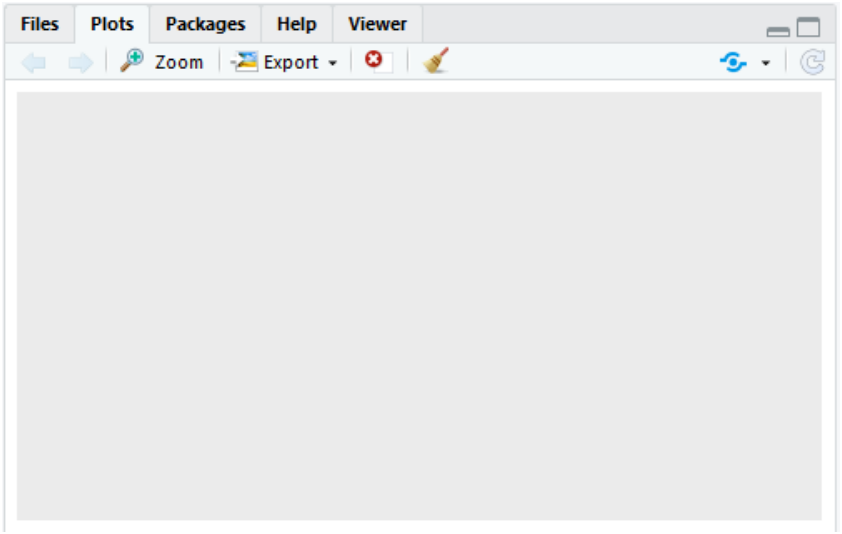
Шевцов Василий Викторович,
директор ДИТ РУДН, shevtsov_vv@rudn.university

Компоненты ggplot

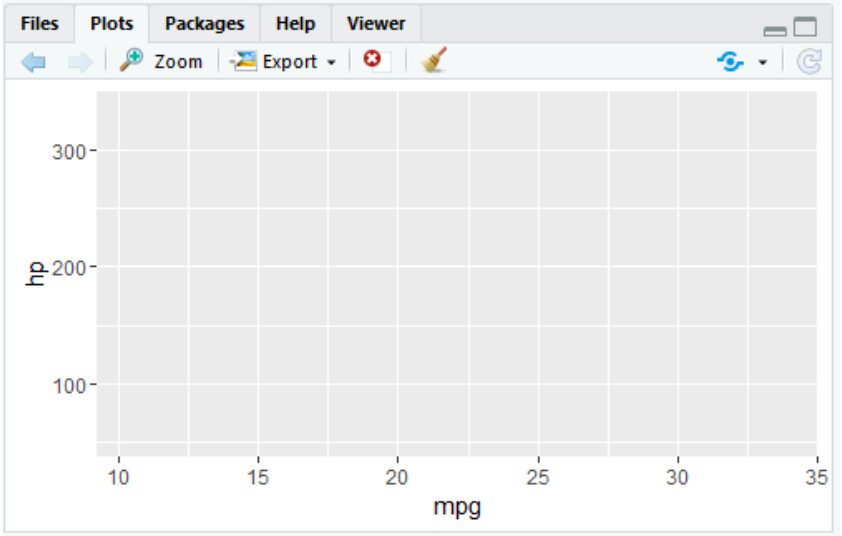
Aesthetic attributes	Определяет данные графика
Geometric objects	Определяет, как будут отображаться данные (вид графика)
Statistical transformations	Определяет трансформации с данными, которые будут отображаться на графике (например, регрессионная прямая или сглаживание)
Scales	Настройка шкал
Coordinates	Настройка системы координат
Faceting	Группировка данных

Aesthetic attributes

ggplot(mtcars)

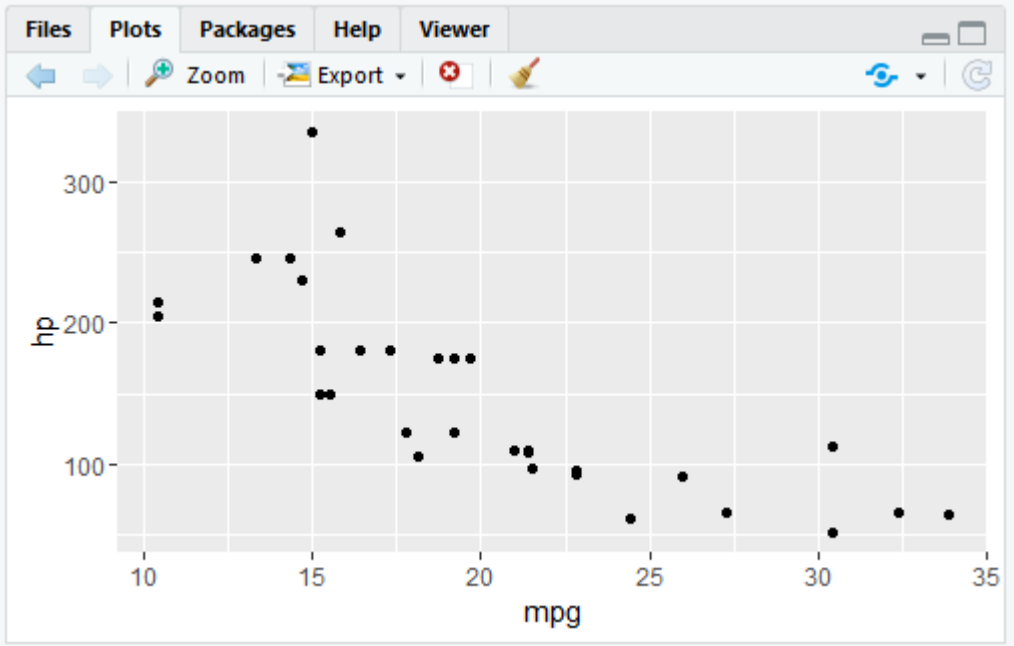


ggplot(mtcars, aes(mpg,hp))

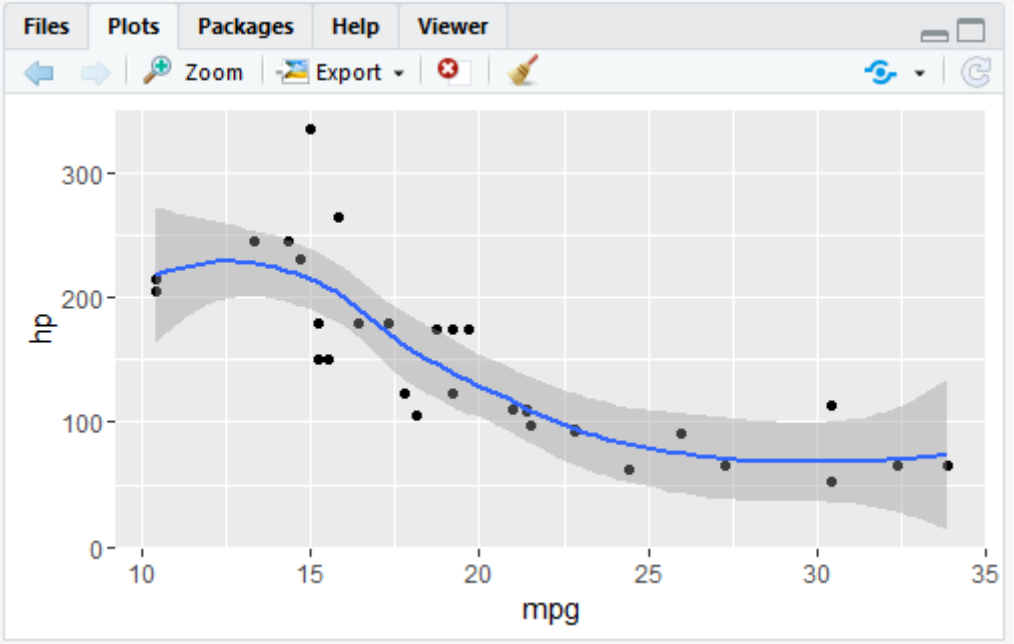


Geometric objects

```
ggplot(mtcars, aes(mpg,hp))+
geom_point()
```

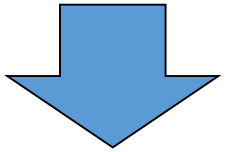



```
ggplot(mtcars, aes(mpg,hp))+
geom_point()+
geom_smooth()
```

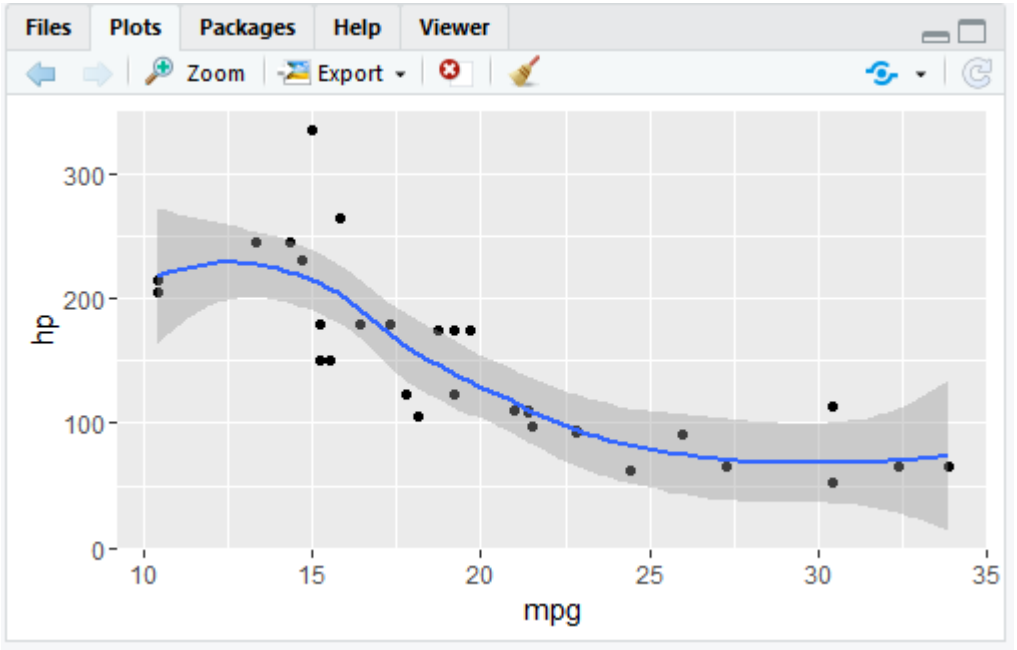


Передача параметров

```
ggplot(mtcars, aes(mpg,hp))+
geom_point()+
geom_smooth()
```

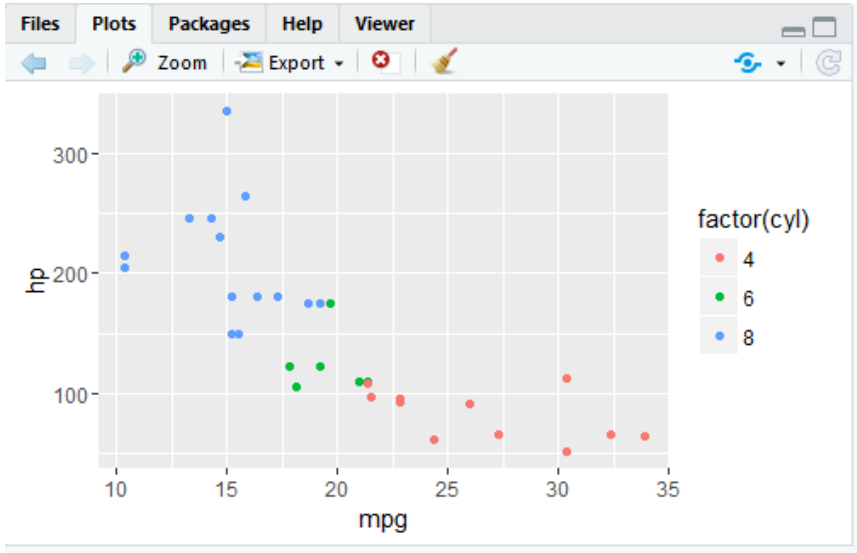


```
ggplot(mtcars)+
geom_point(aes(mpg,hp))+
geom_smooth(aes(mpg,hp))
```

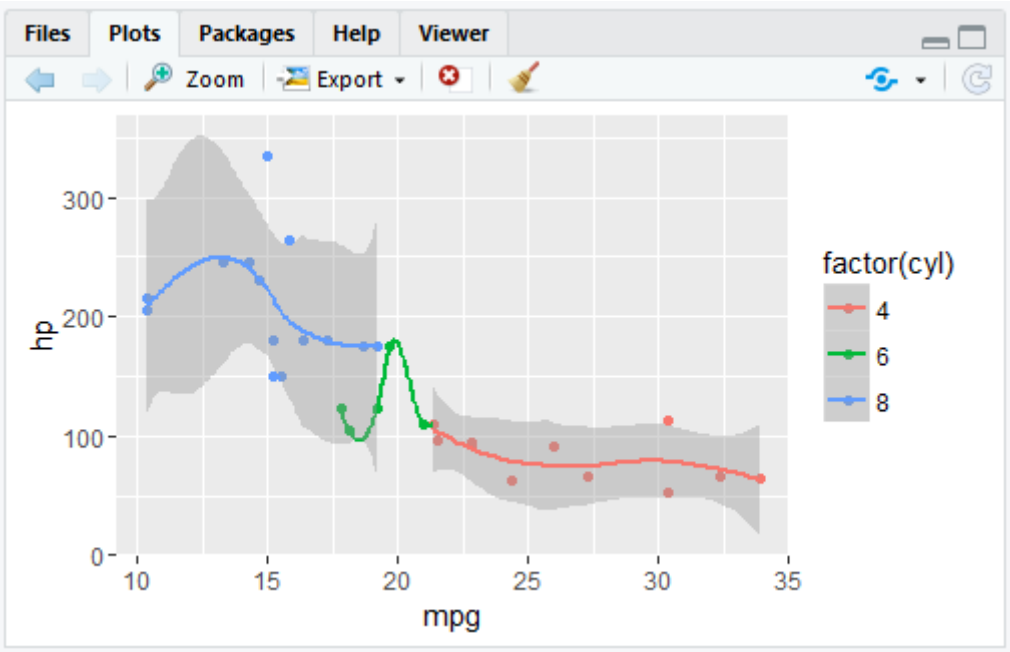


Передача параметров

```
ggplot(mtcars,
aes(mpg,hp,color=factor(cyl))
)+
geom_point()
```

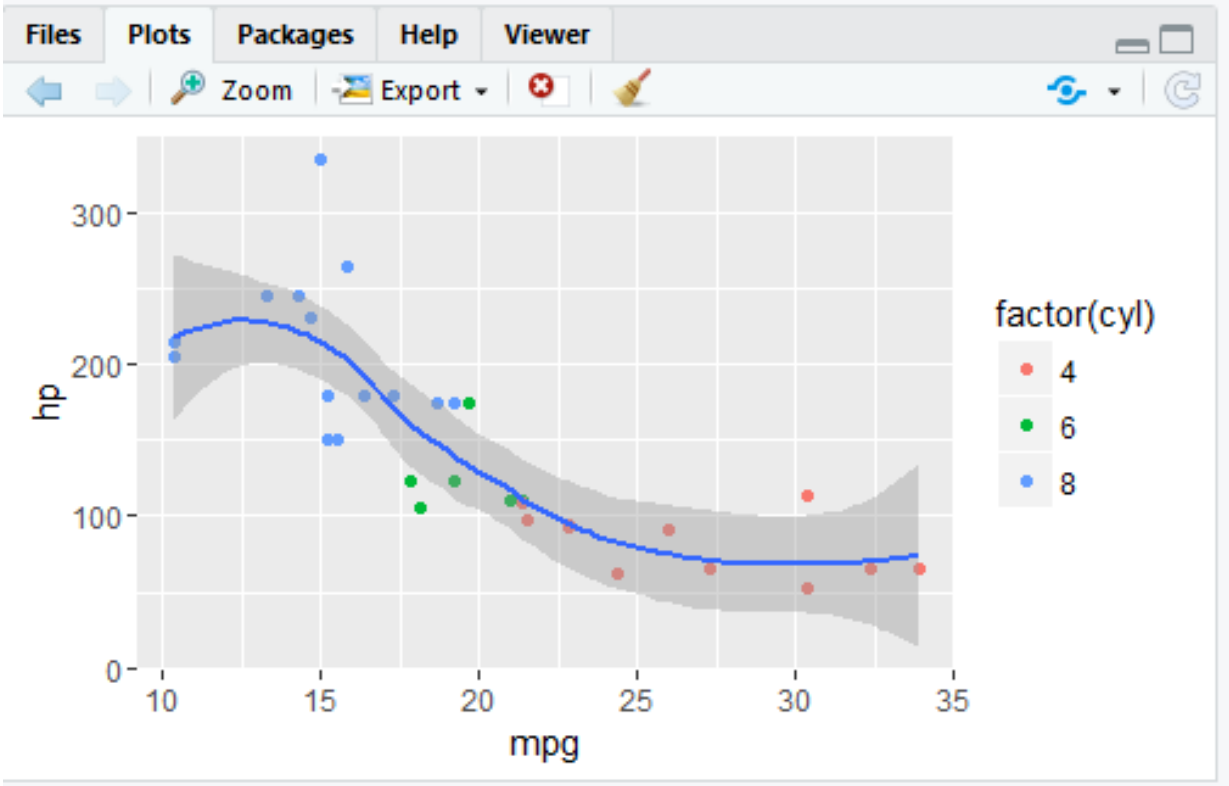



```
ggplot(mtcars,
aes(mpg,hp,color=factor(cyl))
)+
geom_point()+
geom_smooth()
```



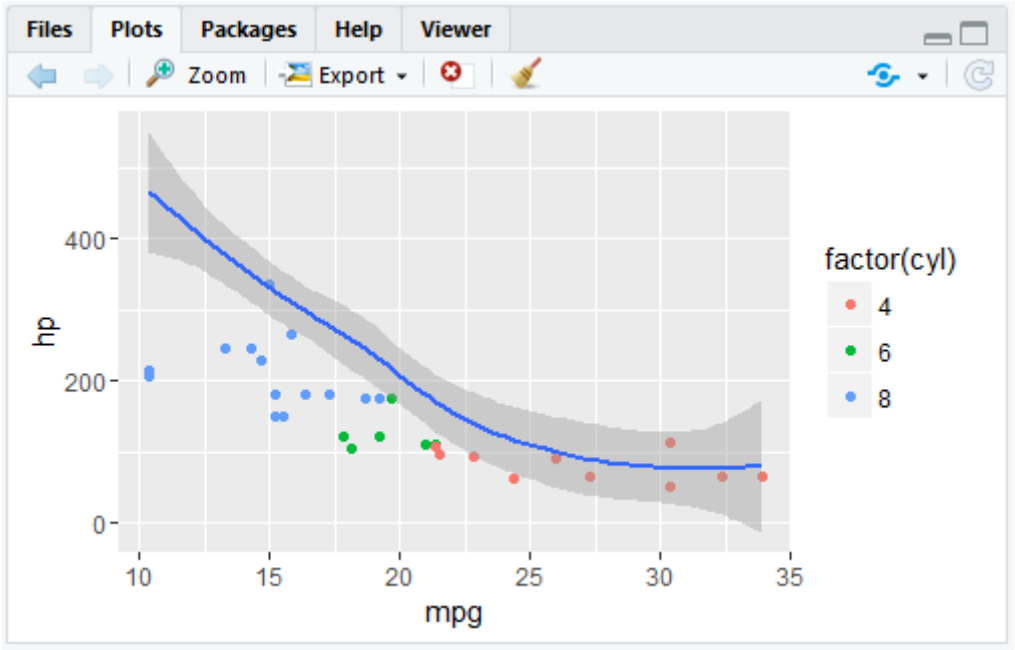
Передача параметров

```
ggplot(mtcars, aes(mpg,hp))+
geom_point(aes(color=factor(cyl)))+
geom_smooth()
```



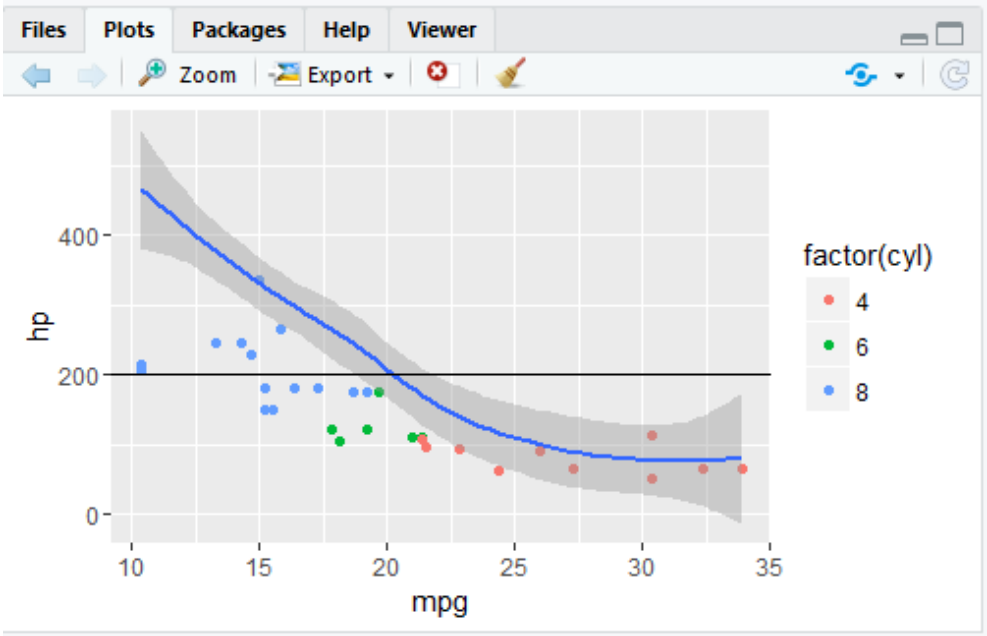
Передача параметров

```
ggplot(mtcars, aes(mpg,hp))+
geom_point(aes(color=factor(cyl)))+
geom_smooth(aes(mpg,disp))
```

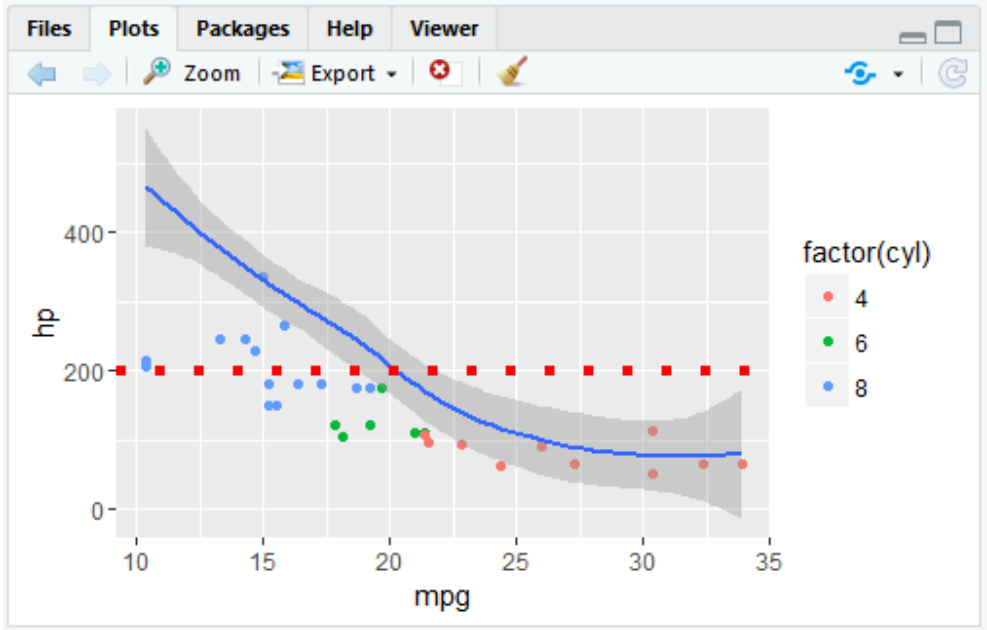


Geom без данных

```
ggplot(mtcars, aes(mpg,hp))+
geom_point(aes(color=factor(cyl)))+
geom_smooth(aes(mpg,disp))+
geom_hline(yintercept = 200)
```



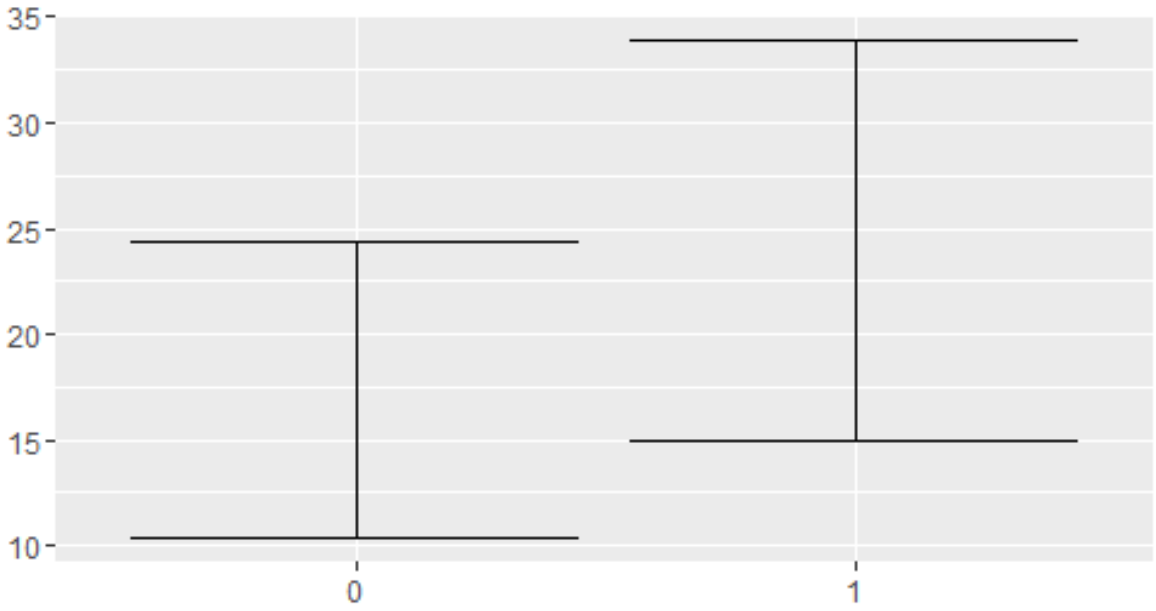

```
ggplot(mtcars, aes(mpg,hp))+
geom_point(aes(color=factor(cyl)))+
geom_smooth(aes(mpg,disp))+
geom_hline(yintercept = 200,
           color="red",
           linetype="dotted",
           size=2)
```



Geom_errorbar

geom_linerange understands the following aesthetics (required aesthetics are in bold):

- **x**
- **ymax**
- **ymin**
- alpha
- colour
- linetype
- size



Geom_errorbar

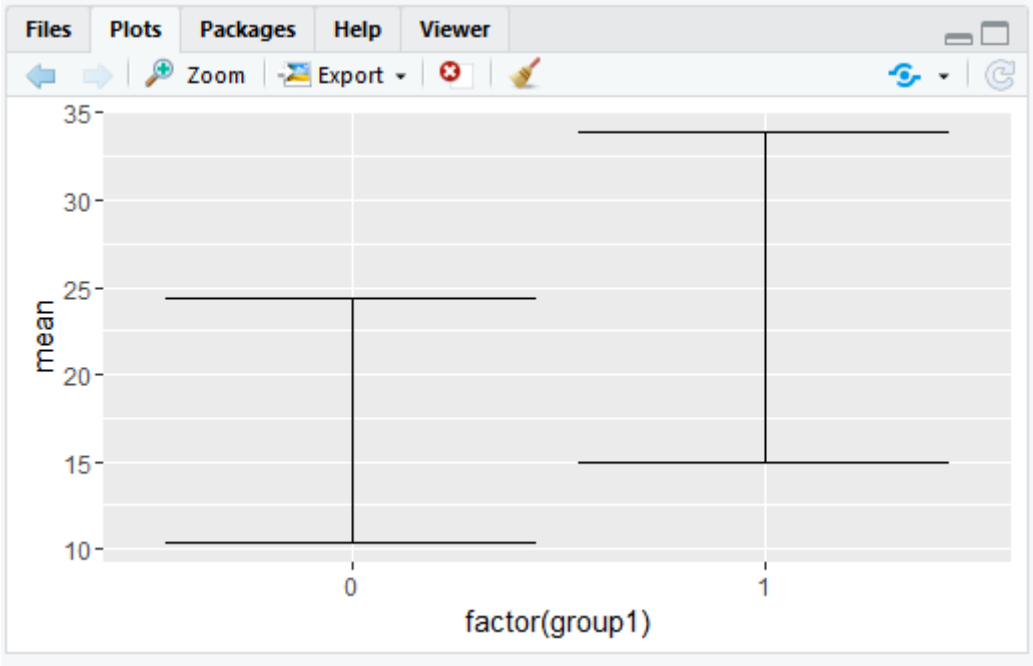
```
install.packages("psych")
library(psych)
df1 <- mtcars

df2 <- describeBy(
            x=df1$mpg,
            group = df1$am,
            mat=TRUE,
            digits=2,
            fast=TRUE)

ggplot(df2,aes(
            x=group1
            ,y=mean))+
             geom_errorbar(aes(
                        ymin=min,
                        ymax=max))
```

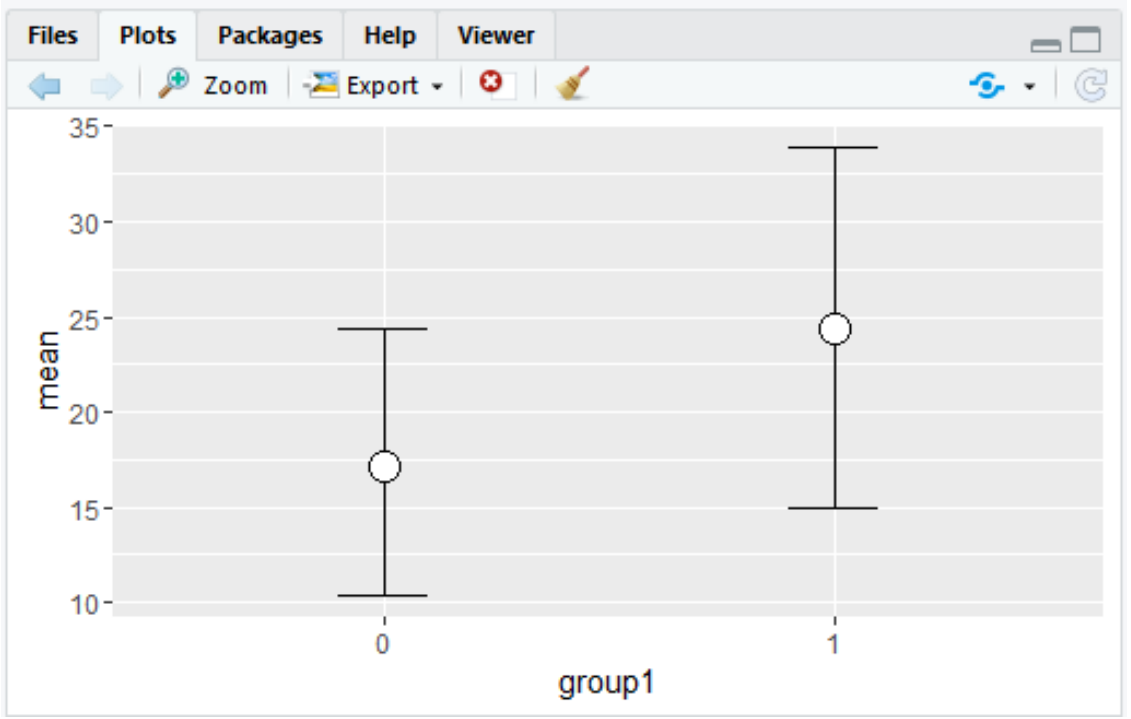
Filter

	item	group1	vars	n	mean	sd	min	max	range	se
X11	1	0	1	19	17.15	3.83	10.4	24.4	14.0	0.88
X12	2	1	1	13	24.39	6.17	15.0	33.9	18.9	1.71



Geom_errorbar

```
ggplot(df2,aes(x=group1,y=mean))+
        geom_errorbar(
                aes(
                ymin=min,
                ymax=max),
                width=0.2)+
        geom_point(
                size=5,
                shape=21,
                fill="white")
```

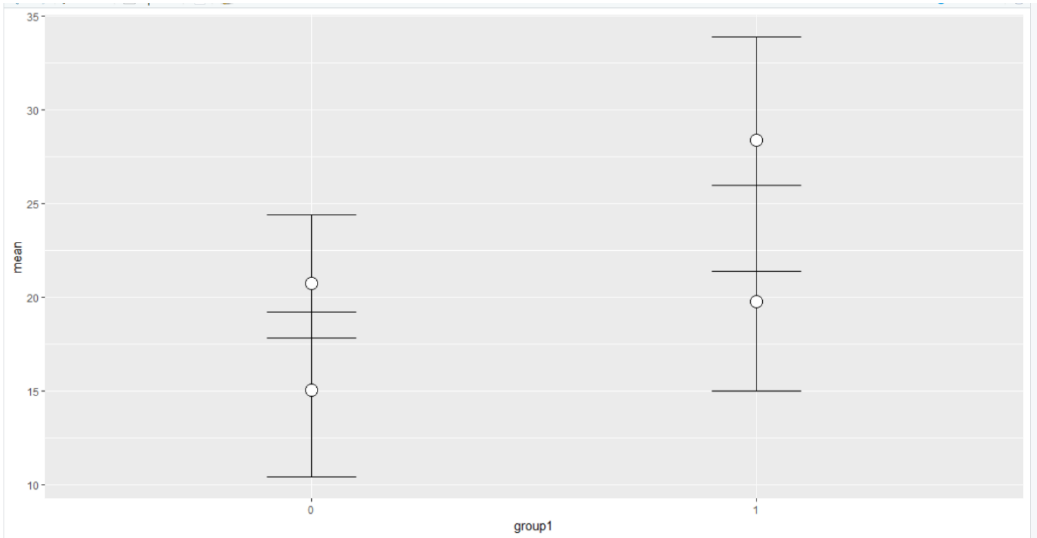


Группировка по двум параметрам

```
library(psych)
df1 <- mtcars
df2 <- describeBy(x=df1$mpg,group = list(df1$am,df1$vs), mat=TRUE,
digits=2, fast=TRUE)
```

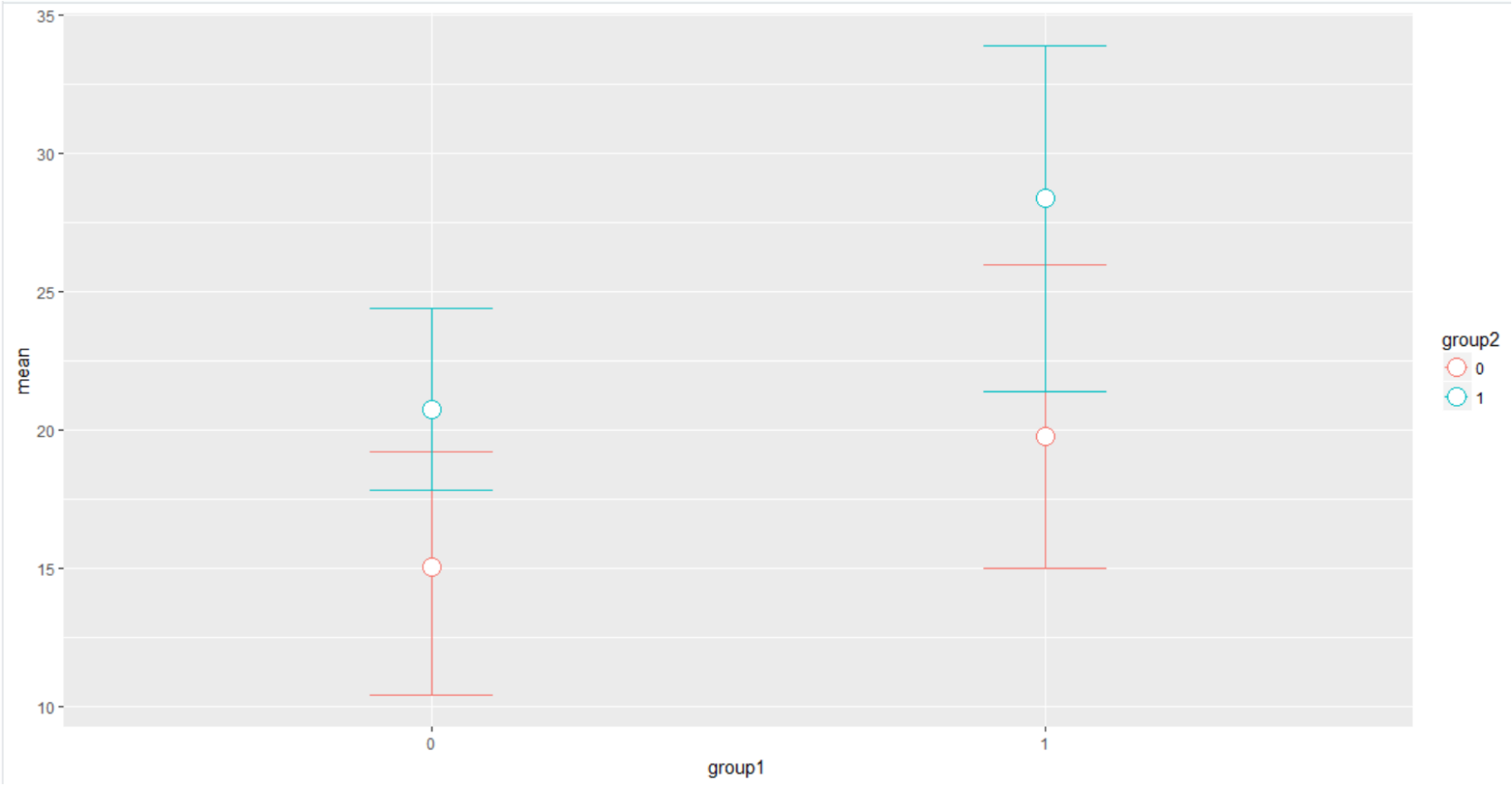
	item	group1	group2	vars	n	mean	sd	min	max	range	se
X11	1	0	0	1	12	15.05	2.77	10.4	19.2	8.8	0.80
X12	2	1	0	1	6	19.75	4.01	15.0	26.0	11.0	1.64
X13	3	0	1	1	7	20.74	2.47	17.8	24.4	6.6	0.93
X14	4	1	1	1	7	28.37	4.76	21.4	33.9	12.5	1.80

```
ggplot(df2,
aes(x=group1,y=mean))+
geom_errorbar(aes(ymin=min,
              ymax=max),
              width=0.2)+
 geom_point(size=5,
              shape=21,
              fill="white")
```



Группировка по двум параметрам

```
ggplot(df2,aes(x=group1,y=mean,color=group2))+
  geom_errorbar(aes(ymin=min, ymax=max),width=0.2)+
  geom_point(size=5,shape=21,fill="white")
```

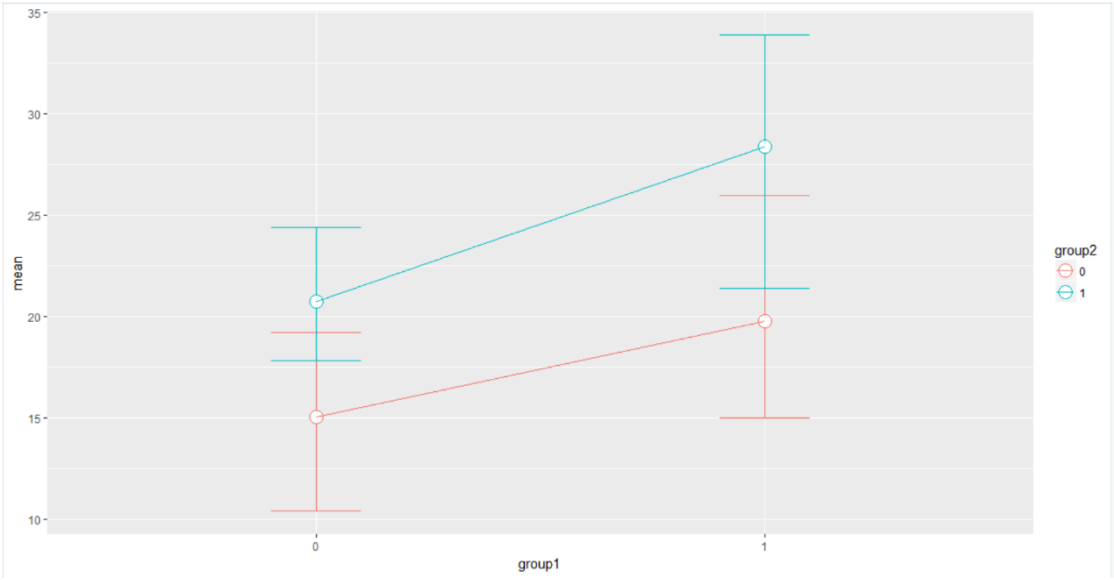


Группировка по двум параметрам

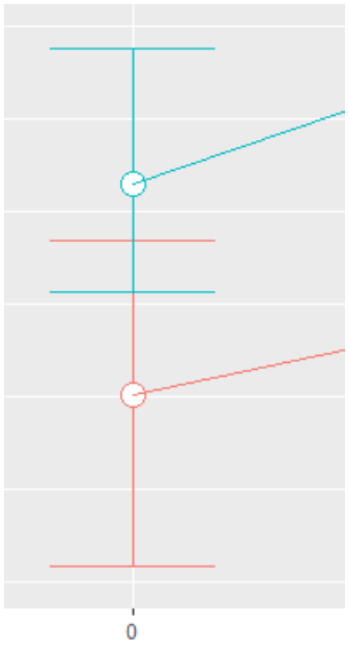
```
ggplot(df2,aes(x=group1,y=mean,color=group2))+
  geom_errorbar(aes(ymin=min, ymax=max),width=0.2)+
  geom_point(size=5,shape=21,fill="white")+
  geom_line()
```

```
geom_path: Each group consists of only one observation. Do you
need to adjust the group aesthetic?
```

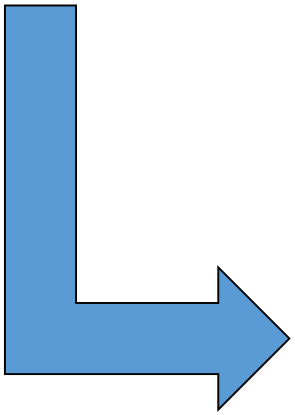
```
ggplot(df2,aes(x=group1,y=mean,color=group2,group=group2))+
  geom_errorbar(aes(ymin=min, ymax=max),width=0.2)+
  geom_point(size=5,shape=21,fill="white")+
  geom_line()
```

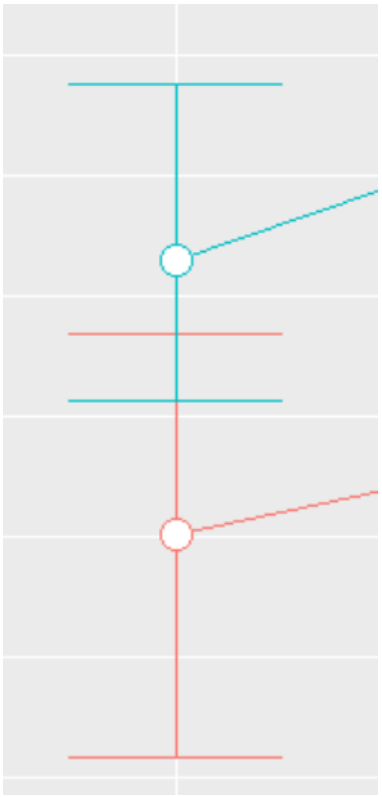


Порядок нанесения слоев



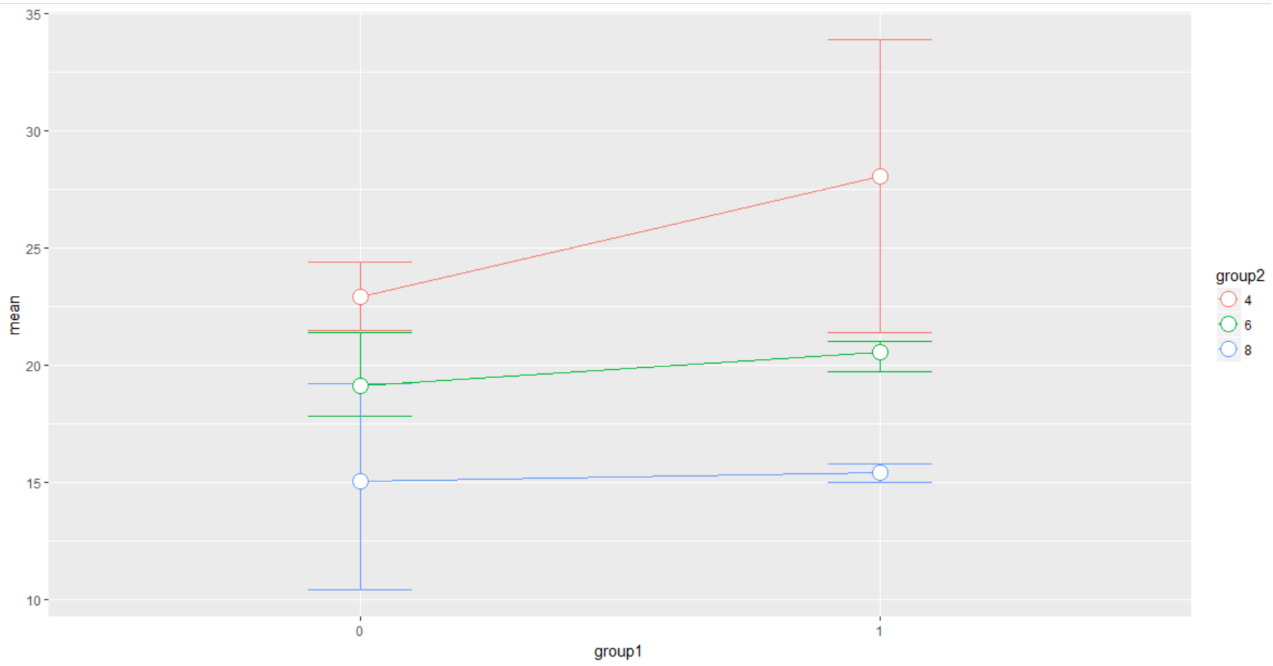

```
ggplot(df2,aes(x=group1,y=mean,color=group2,
group=group2))+
  geom_errorbar(aes(ymin=min, ymax=max),width=0.2)+
  geom_line()+
  geom_point(size=5,shape=21,fill="white")
```





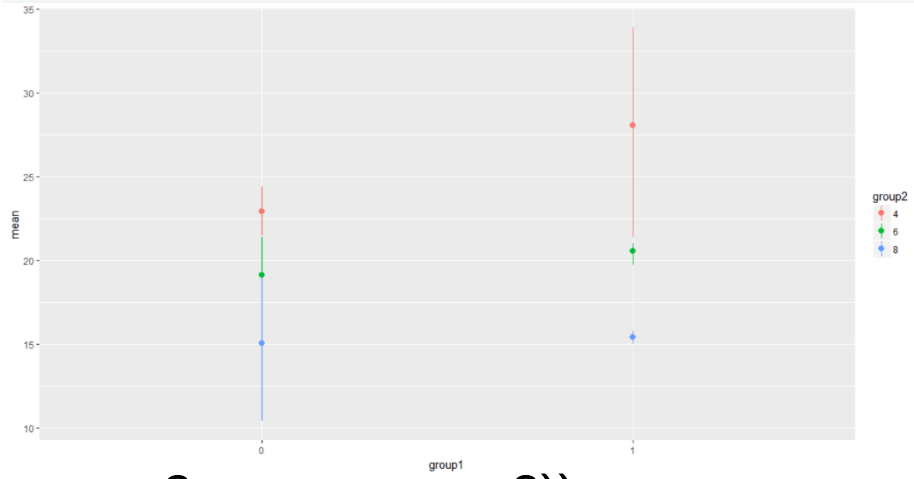
Группировка по трем параметрам

```
library(psych)
df1 <- mtcars
df2 <- describeBy(x=df1$mpg,group = list(df1$am,df1$cyl), mat=TRUE,
digits=2, fast=TRUE)
ggplot(df2,aes(x=group1,y=mean,color=group2,group=group2))+
  geom_errorbar(aes(ymin=min, ymax=max),width=0.2)+
  geom_line()+
  geom_point(size=5,shape=21,fill="white")
```

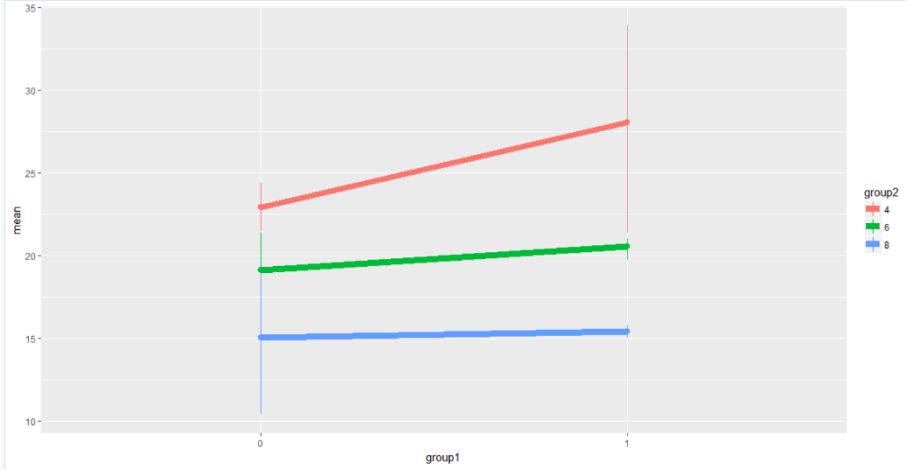


geom_pointrange

```
ggplot(df2,aes(x=group1,y=mean,color=group2,group=group2))+
  geom_pointrange(aes(ymin=min, ymax=max))
```

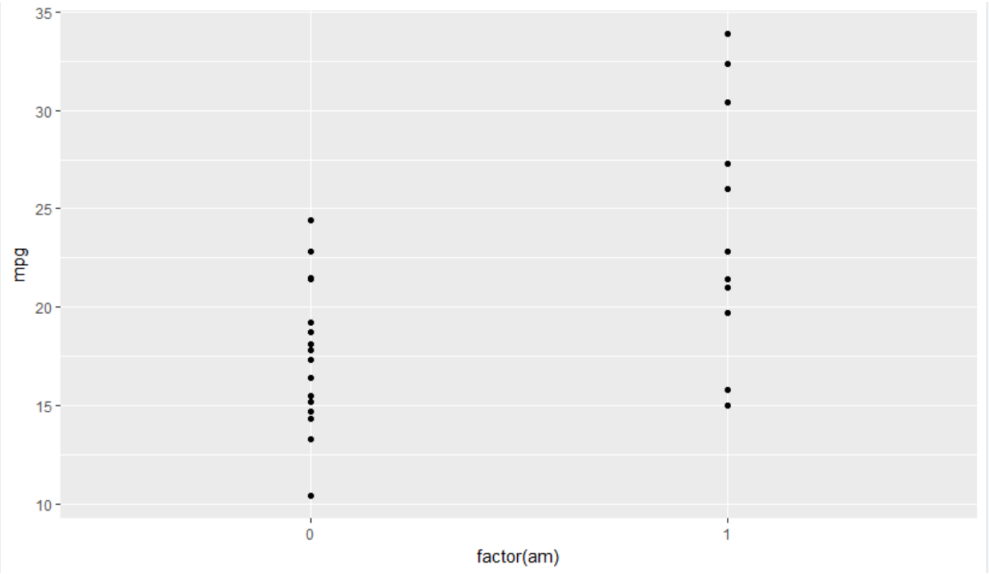



```
ggplot(df2,aes(x=group1,y=mean,color=group2,group=group2))+
  geom_pointrange(aes(ymin=min, ymax=max))+
  geom_line(size=3)
```



Перенос обработки данных в ggplot()

```
df1 <- mtcars
ggplot(df1,aes(factor(am),mpg))+
  geom_point()
```



?stat_summary

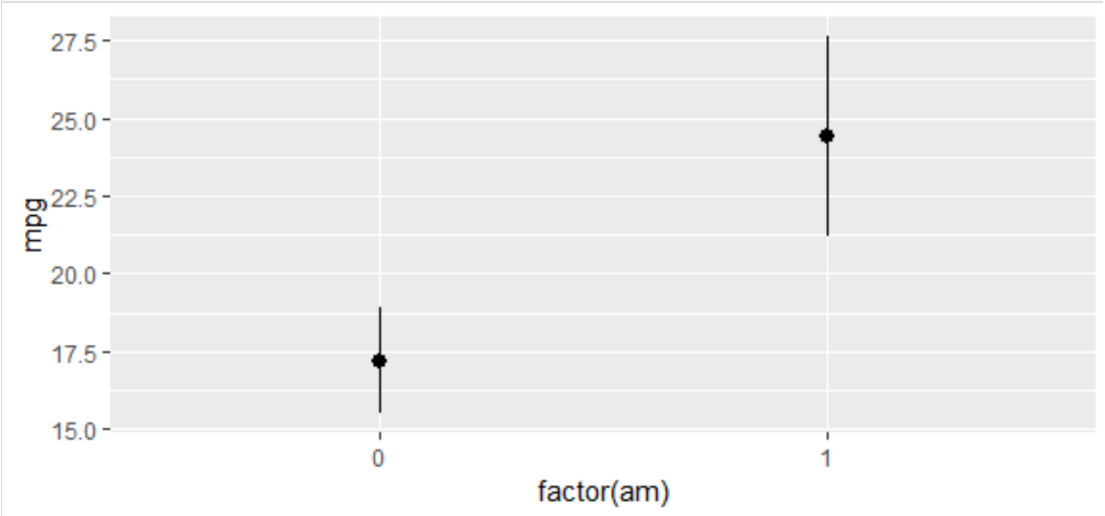
stat_summary(mapping = NULL, data = NULL, geom = "pointrange",
position = "identity", ..., **fun.data** = NULL, fun.y = NULL,
fun.ymax = NULL, fun.ymin = NULL, fun.args = list(), na.rm = FALSE,
show.legend = NA, inherit.aes = TRUE)

fun.data
A function that is given the complete data and should return a data frame with variables ymin, y, and ymax.

Перенос обработки данных в ggplot()

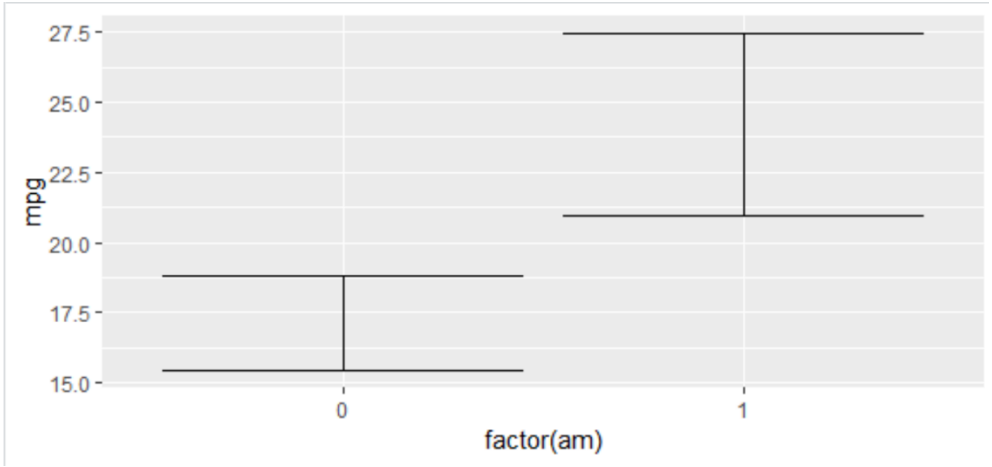
```
> mean_cl_boot(df1$mpg)
         y     ymin     ymax
1 20.09062 18.19352 22.03836
> mean_cl_normal(df1$mpg)
         y     ymin     ymax
1 20.09062 17.91768 22.26357
```

```
ggplot(df1,aes(factor(am),mpg))+
 stat_summary(
          fun.data=mean_cl_boot)
#передается только название
функции
#по умолчанию строится
geom_pointrange
```

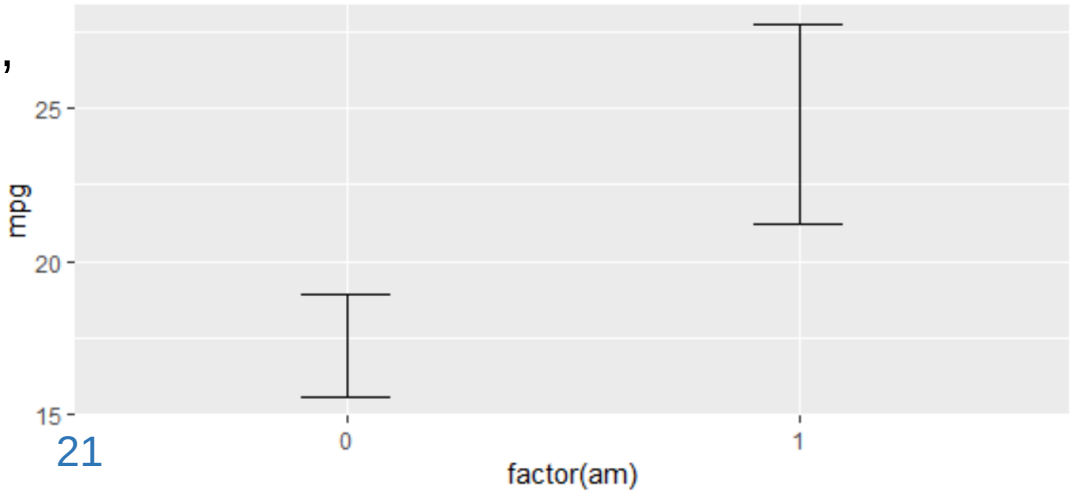


Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar")
```

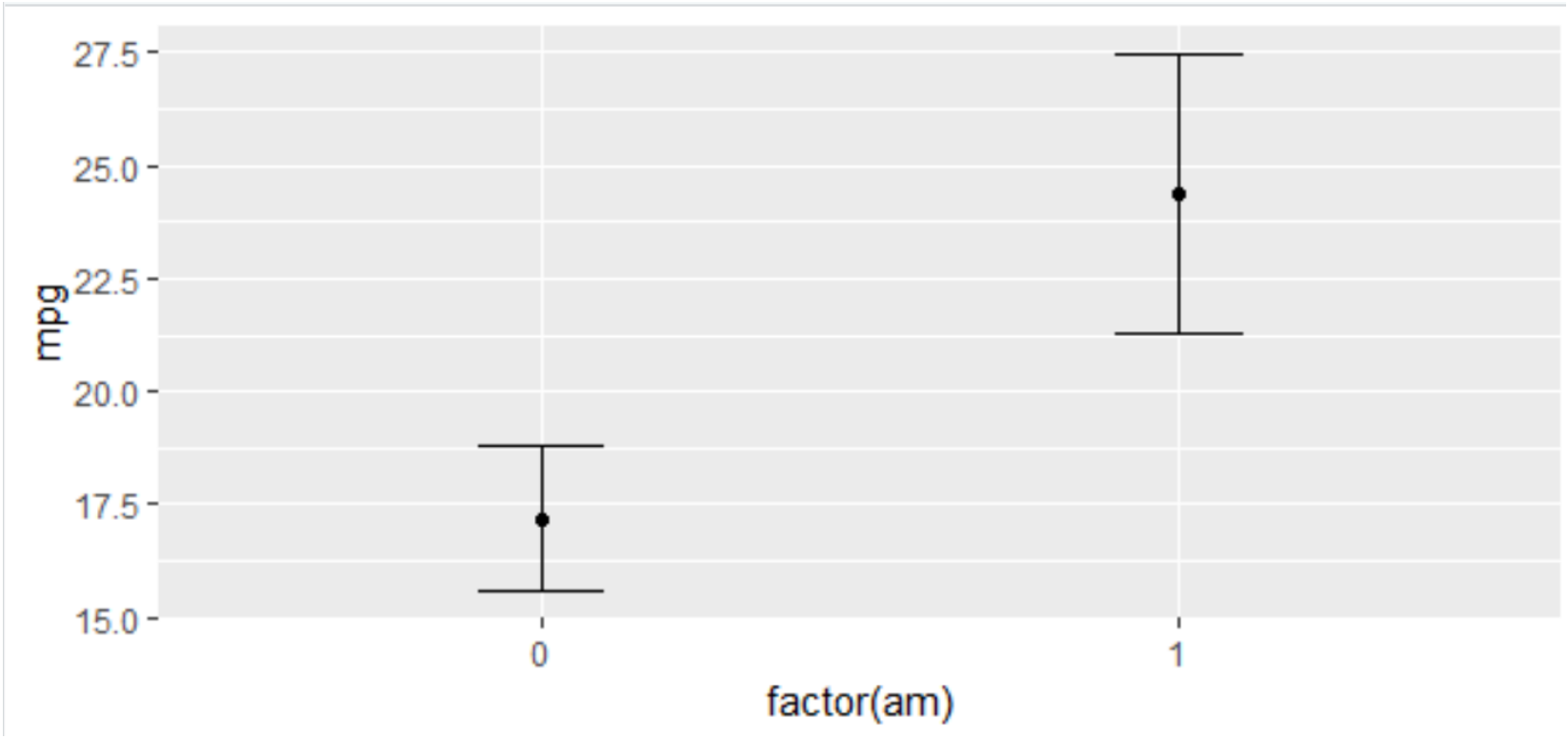



```
ggplot(df1,aes(factor(am),mpg))+
        stat_summary(
        fun.data=mean_cl_boot,
        geom="errorbar",
        width=0.2)
```



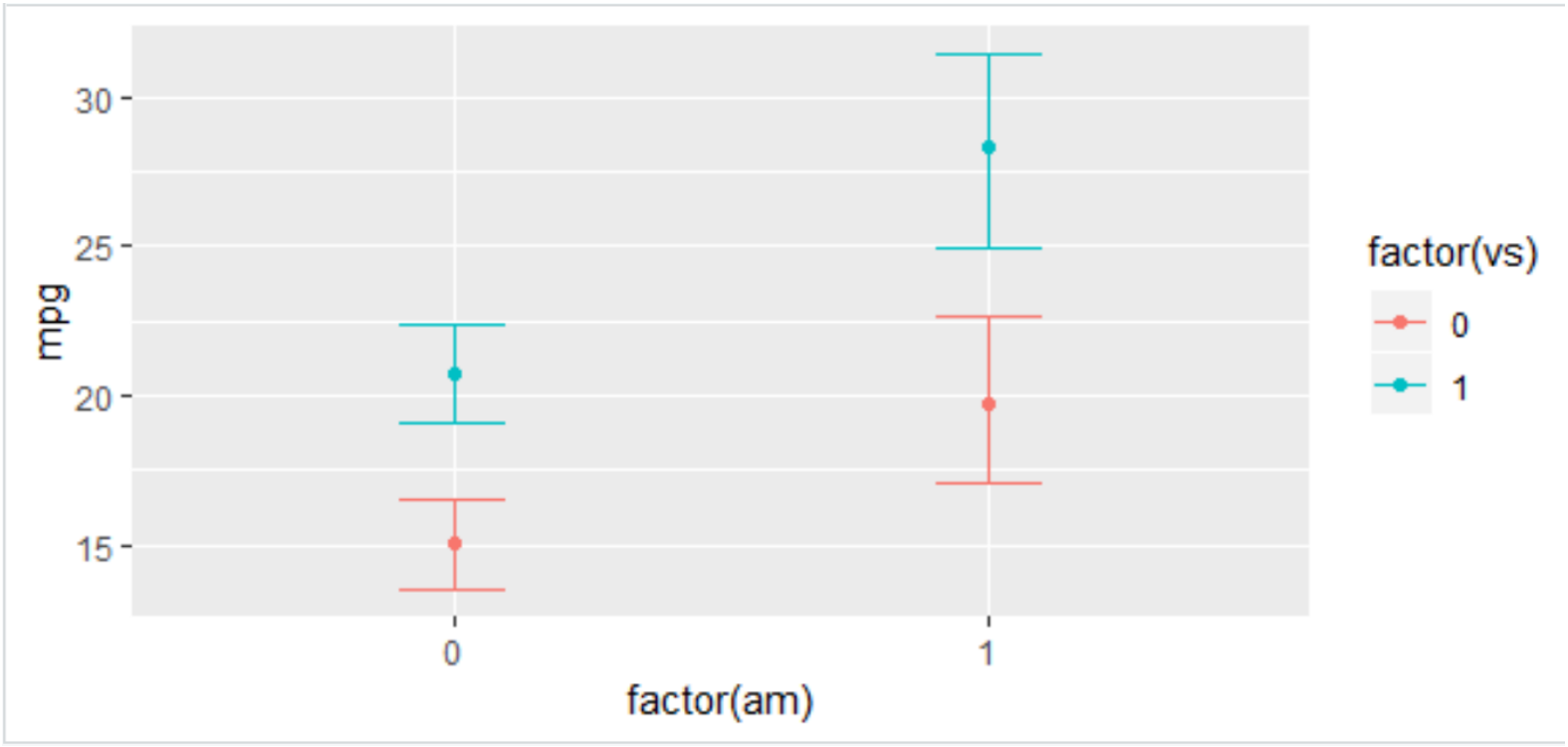
Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar",width=0.2)+
  stat_summary(fun.data=mean_cl_boot,geom="point")
```



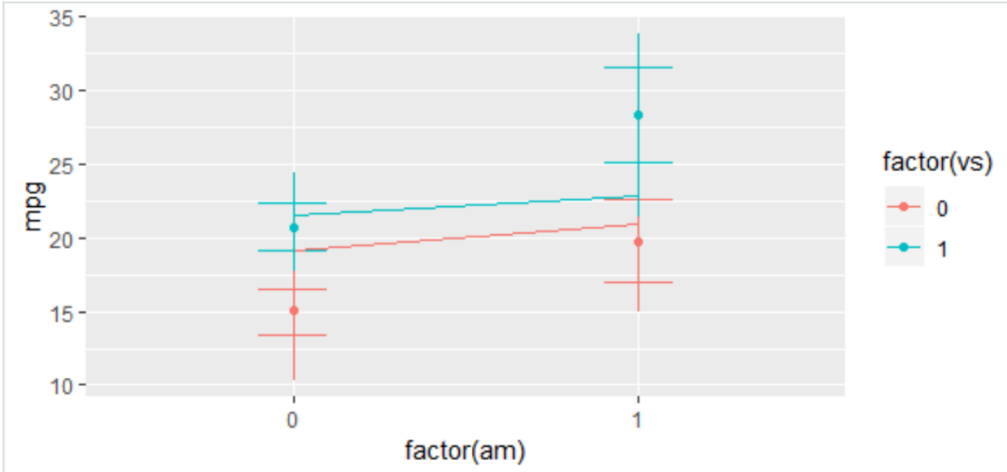
Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg,color=factor(vs)))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar",width=0.2)+
  stat_summary(fun.data=mean_cl_boot,geom="point")
```



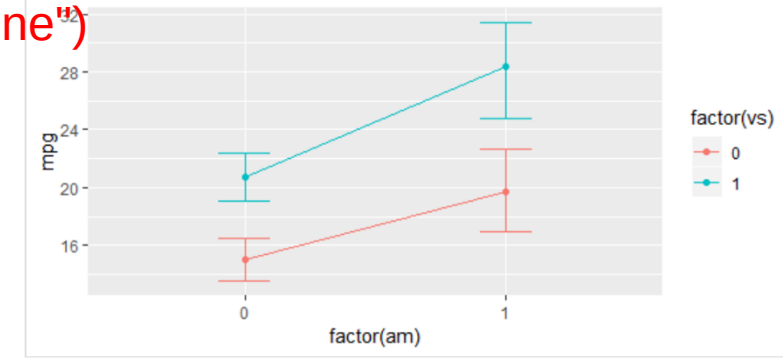
Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg,color=factor(vs),group=factor(vs)))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar",width=0.2)+
  stat_summary(fun.data=mean_cl_boot,geom="point")+
  geom_line()
```



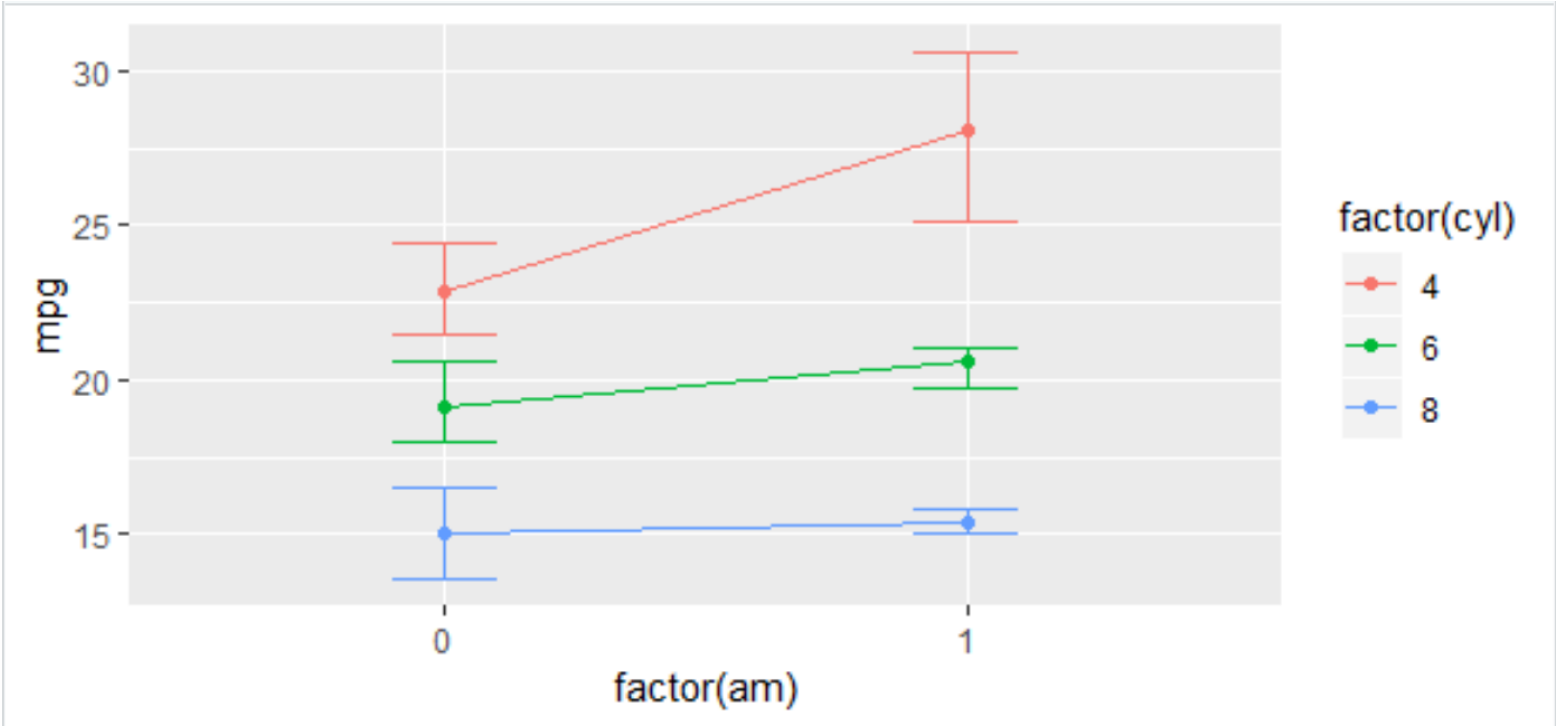

```
ggplot(df1,aes(factor(am),mpg,color=factor(vs),group=factor(vs)))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar",width=0.2)+
  stat_summary(fun.data=mean_cl_boot,geom="point")+
  stat_summary(fun.data=mean_cl_boot,geom="line")
```

```
stat_summary(
        fun.y=mean,geom="line")
```



Перенос обработки данных в ggplot()

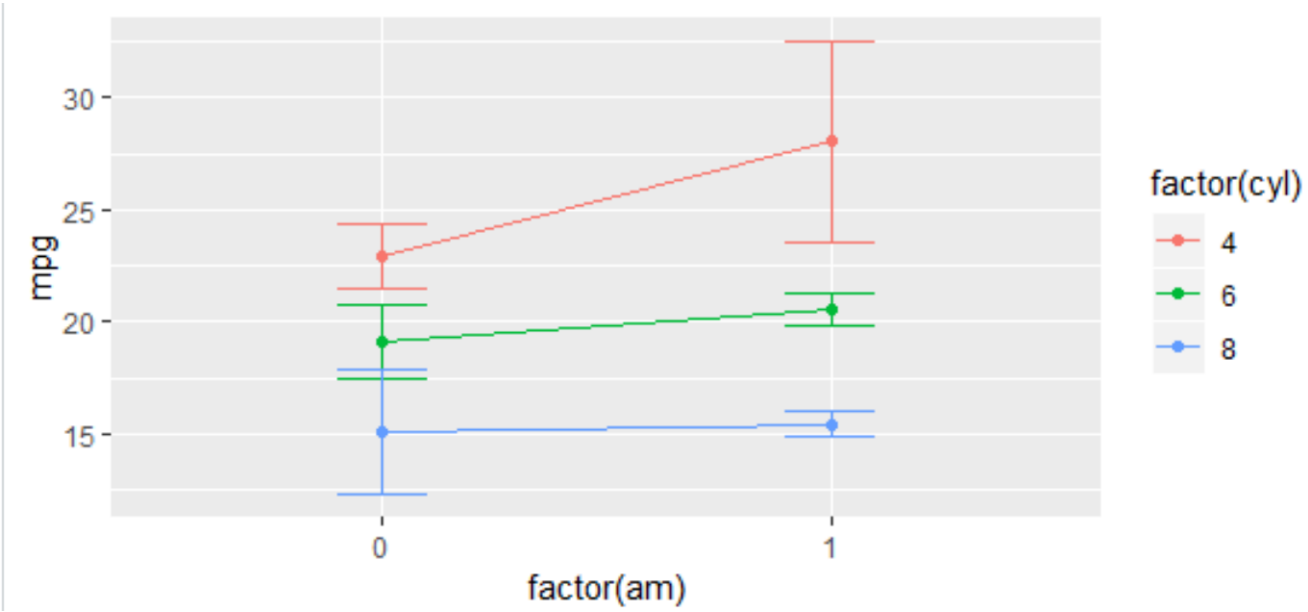
```
ggplot(df1,aes(factor(am),mpg,color=factor(cyl),group=factor(cyl)))+
  stat_summary(fun.data=mean_cl_boot,geom="errorbar",width=0.2)+
  stat_summary(fun.data=mean_cl_boot,geom="point")+
  stat_summary(fun.y=mean,geom="line")
```



Перенос обработки данных в ggplot()

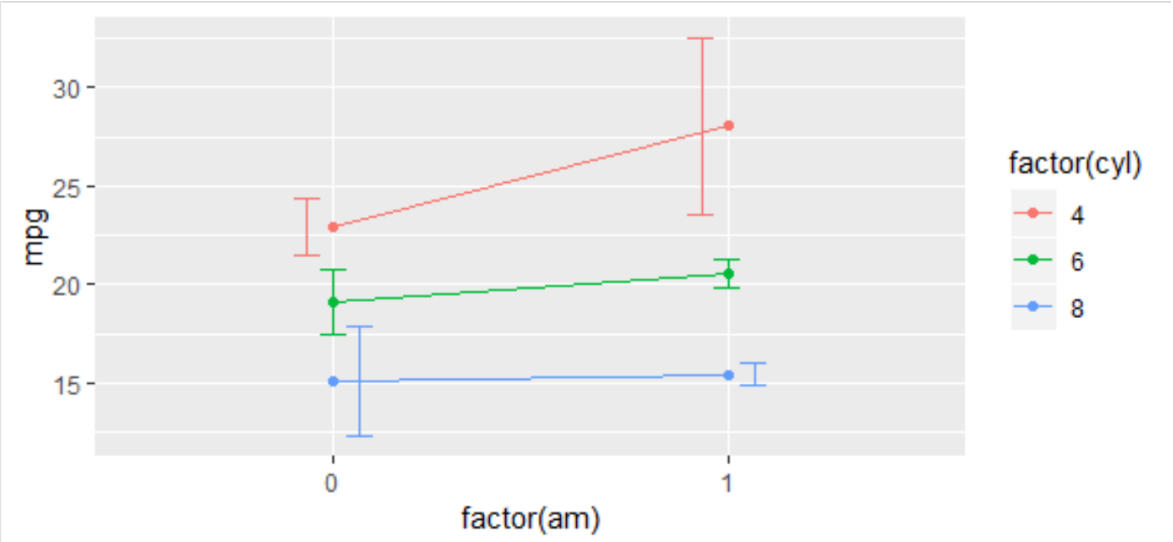
```
sd_err <- function(x){
          c(y=mean(x),ymin=mean(x)-sd(x),ymax=mean(x)+sd(x))}

ggplot(df1,aes(factor(am),mpg,color=factor(cyl),group=factor(cyl)))+
  stat_summary(fun.data=sd_err,geom="errorbar",width=0.2)+
  stat_summary(fun.data=sd_err,geom="point")+
  stat_summary(fun.y=mean,geom="line")
```



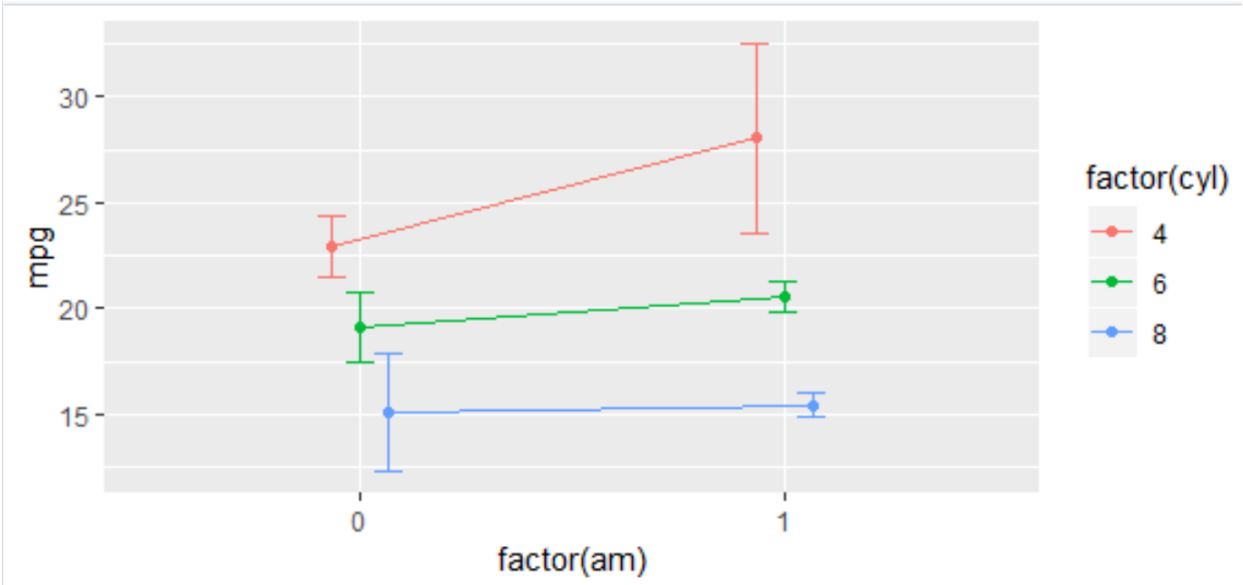
Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg,color=factor(cyl),group=factor(cyl)))+
  stat_summary(
          fun.data=sd_err,
          geom="errorbar",
          width=0.2,
          position = position_dodge(0.2))+
  stat_summary(fun.data=sd_err,geom="point")+
  stat_summary(fun.y=mean,geom="line")
```



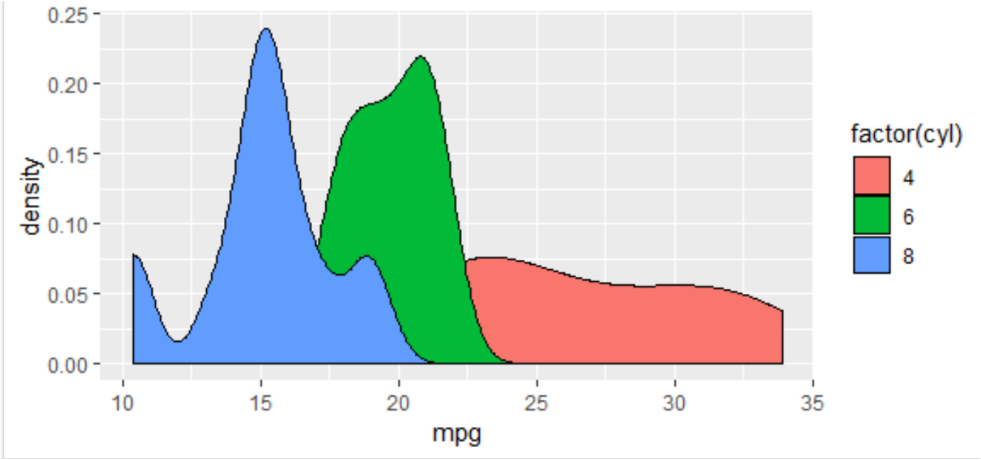
Перенос обработки данных в ggplot()

```
ggplot(df1,aes(factor(am),mpg,color=factor(cyl),group=factor(cyl)))+
  stat_summary(fun.data=sd_err,geom="errorbar",width=0.2,
               position = position_dodge(0.2))+
  stat_summary(fun.data=sd_err,geom="point",
               position = position_dodge(0.2))+
  stat_summary(fun.y=mean,geom="line",
               position = position_dodge(0.2))
```

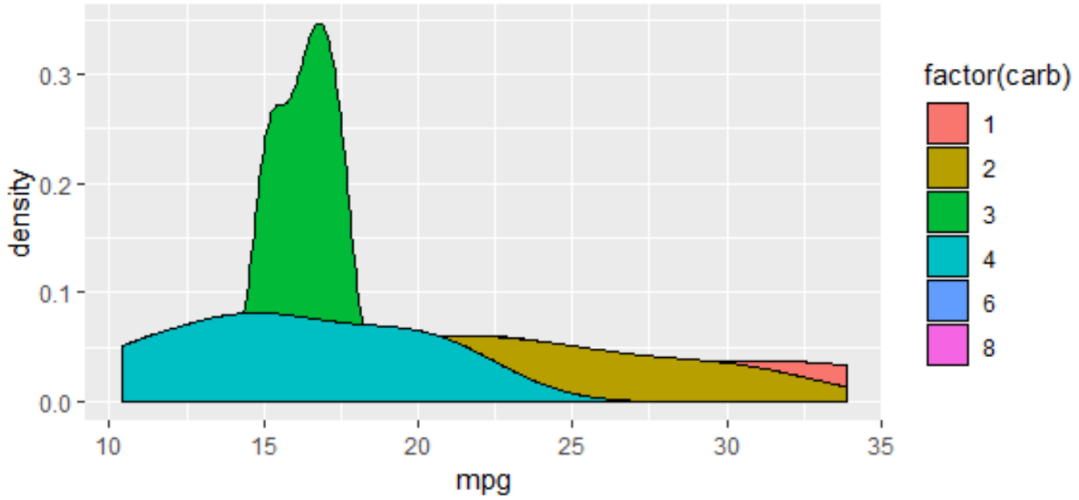


Facet

```
ggplot(df1,aes(mpg,fill=factor(cyl)))+
  geom_density()
```

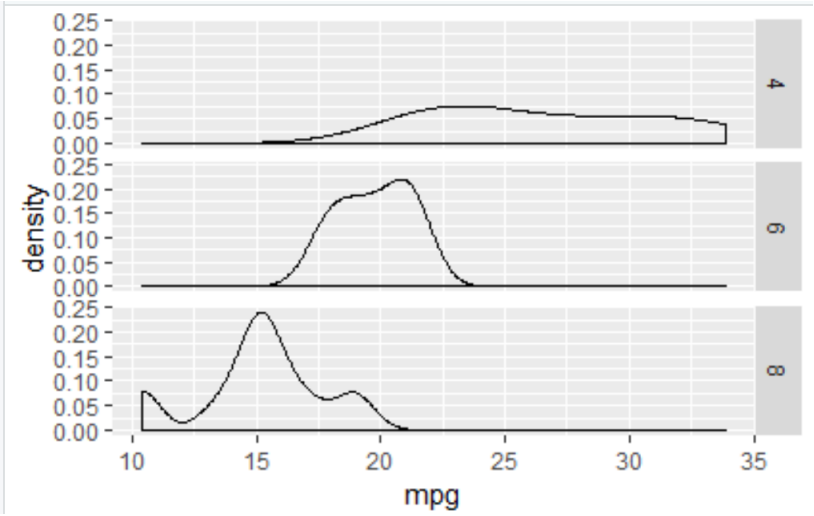



```
ggplot(df1,aes(mpg,fill=factor(carb)))+
  geom_density()
```

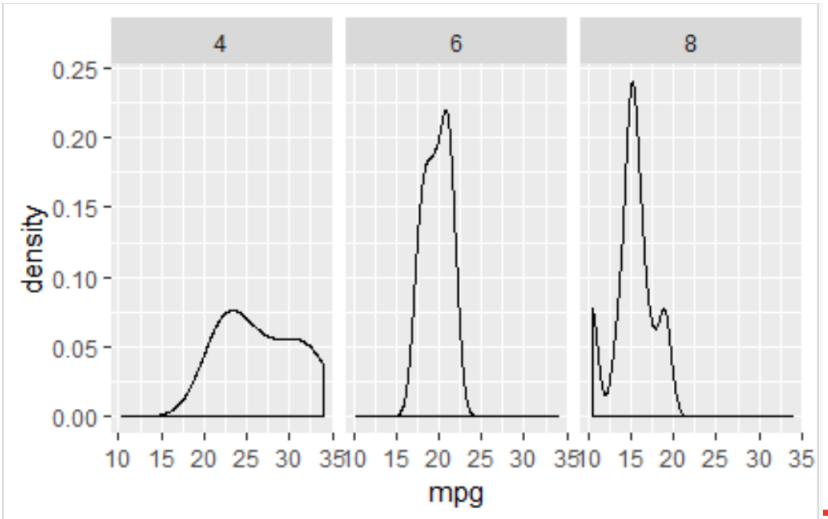


Facet

```
ggplot(df1,aes(mpg))+
  geom_density()+
  facet_grid(factor(cyl)~.)
```

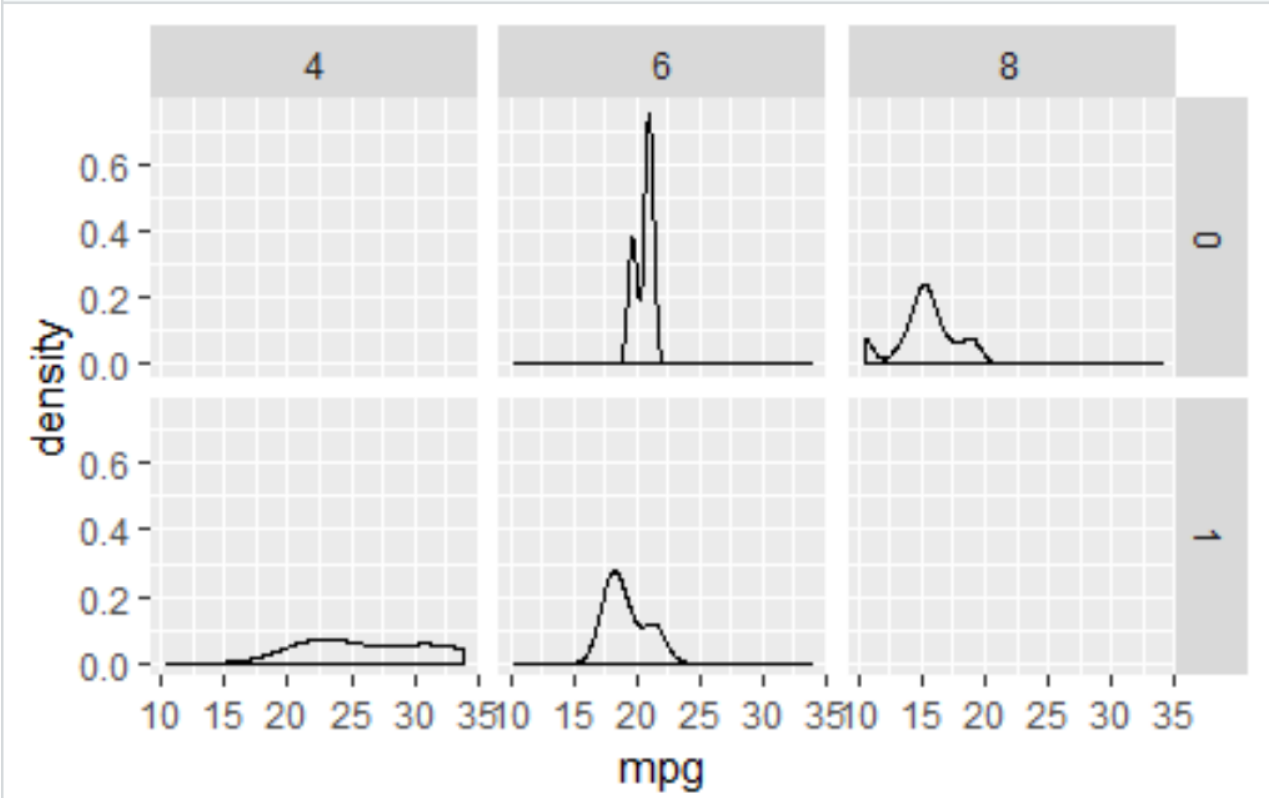



```
ggplot(df1,aes(mpg))+
  geom_density()+
  facet_grid(.~factor(cyl))
```



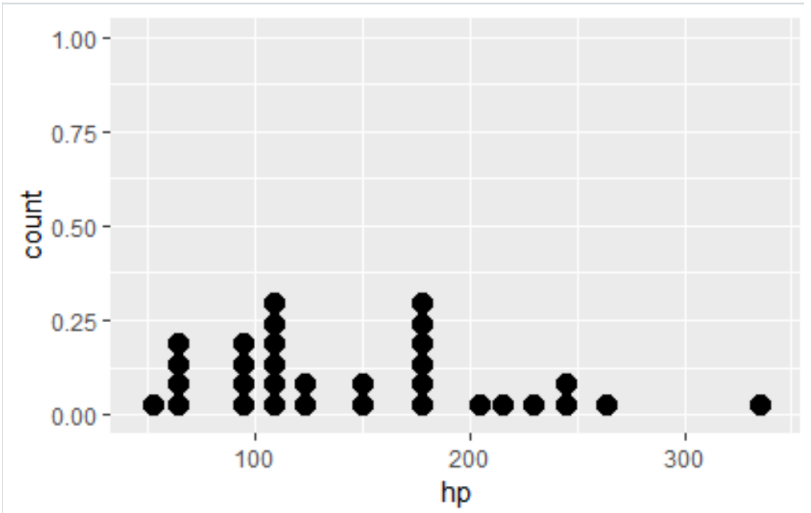
Facet

```
ggplot(df1,aes(mpg))+
  geom_density()+
  facet_grid(factor(vs)~factor(cyl))
```

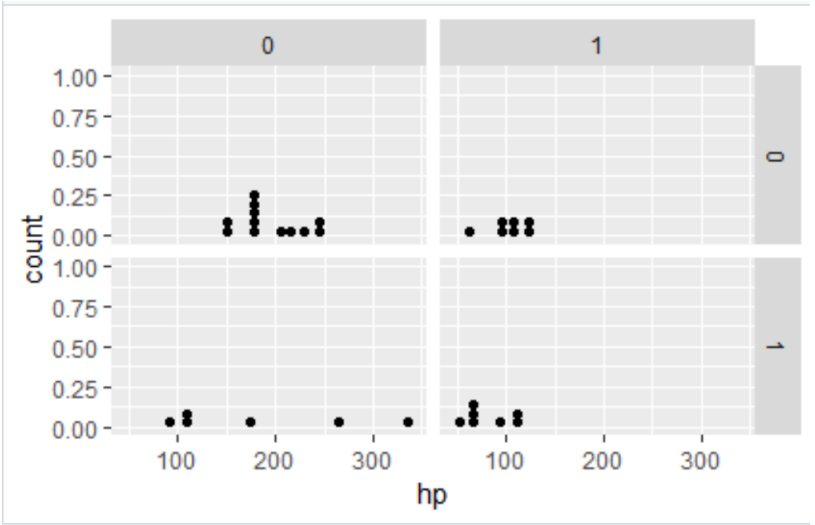


Facet

```
ggplot(df1,aes(hp))+
  geom_dotplot()
```

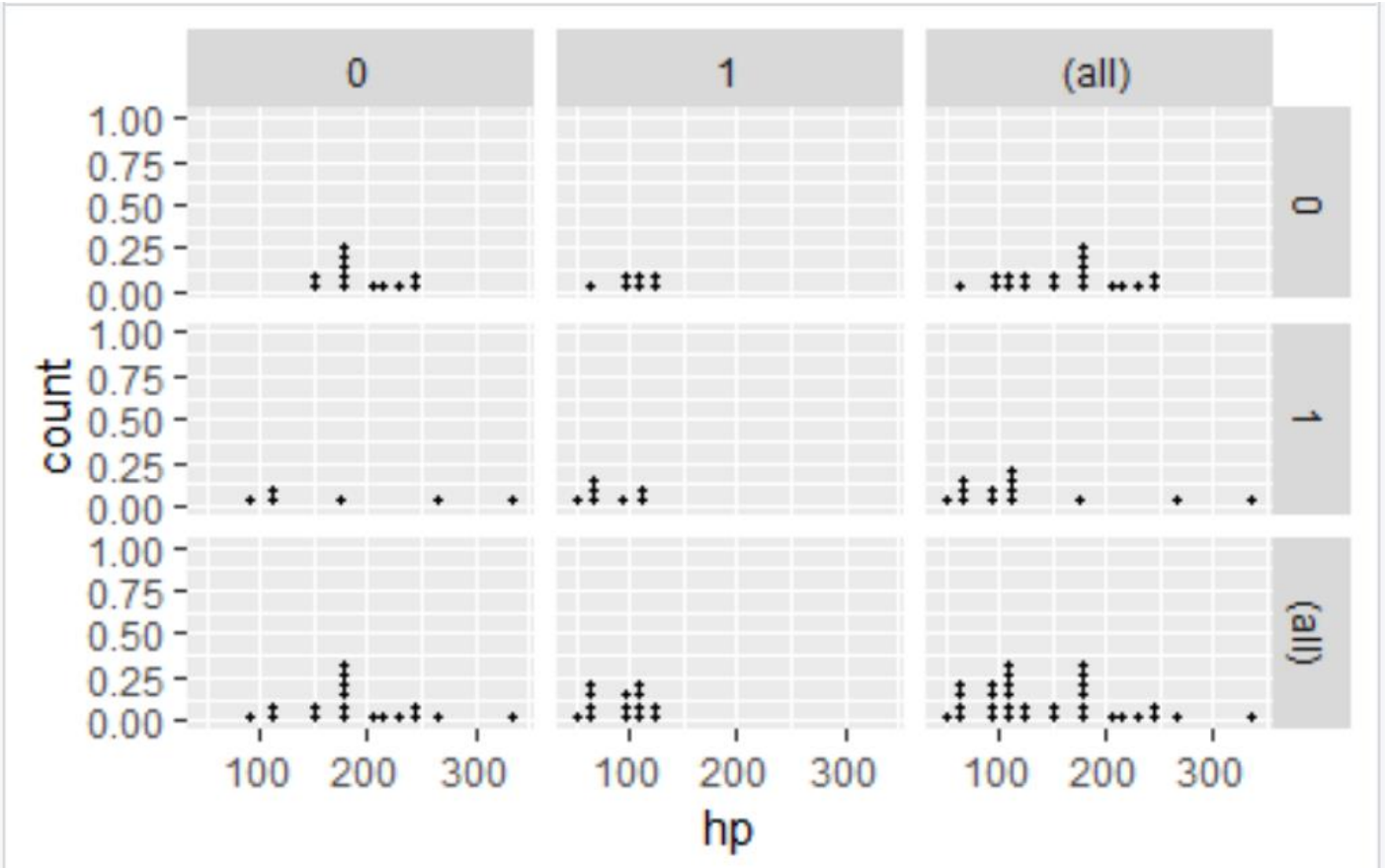



```
ggplot(df1,aes(hp))+
  geom_dotplot()+
  facet_grid(am~vs)
```



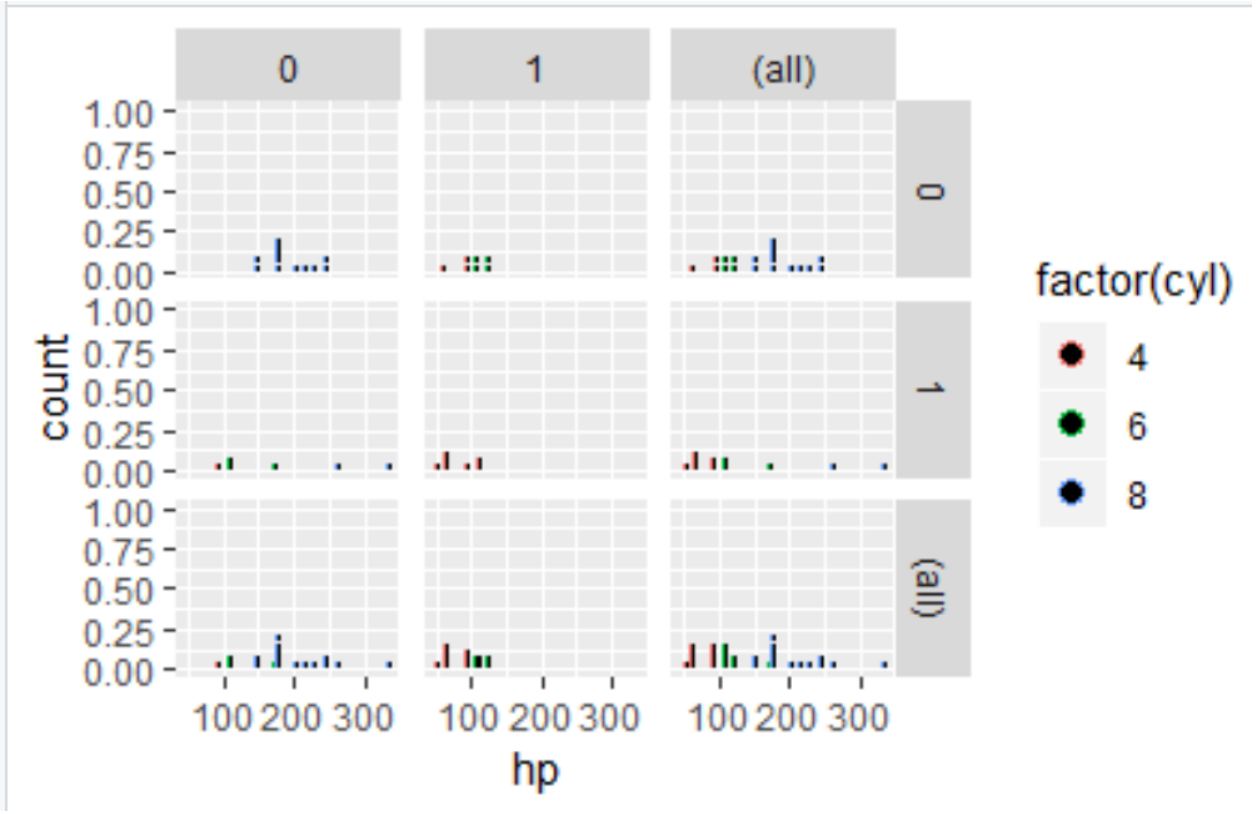
Facet

```
ggplot(df1,aes(hp))+
  geom_dotplot()+
  facet_grid(am~vs,margins = TRUE)
```



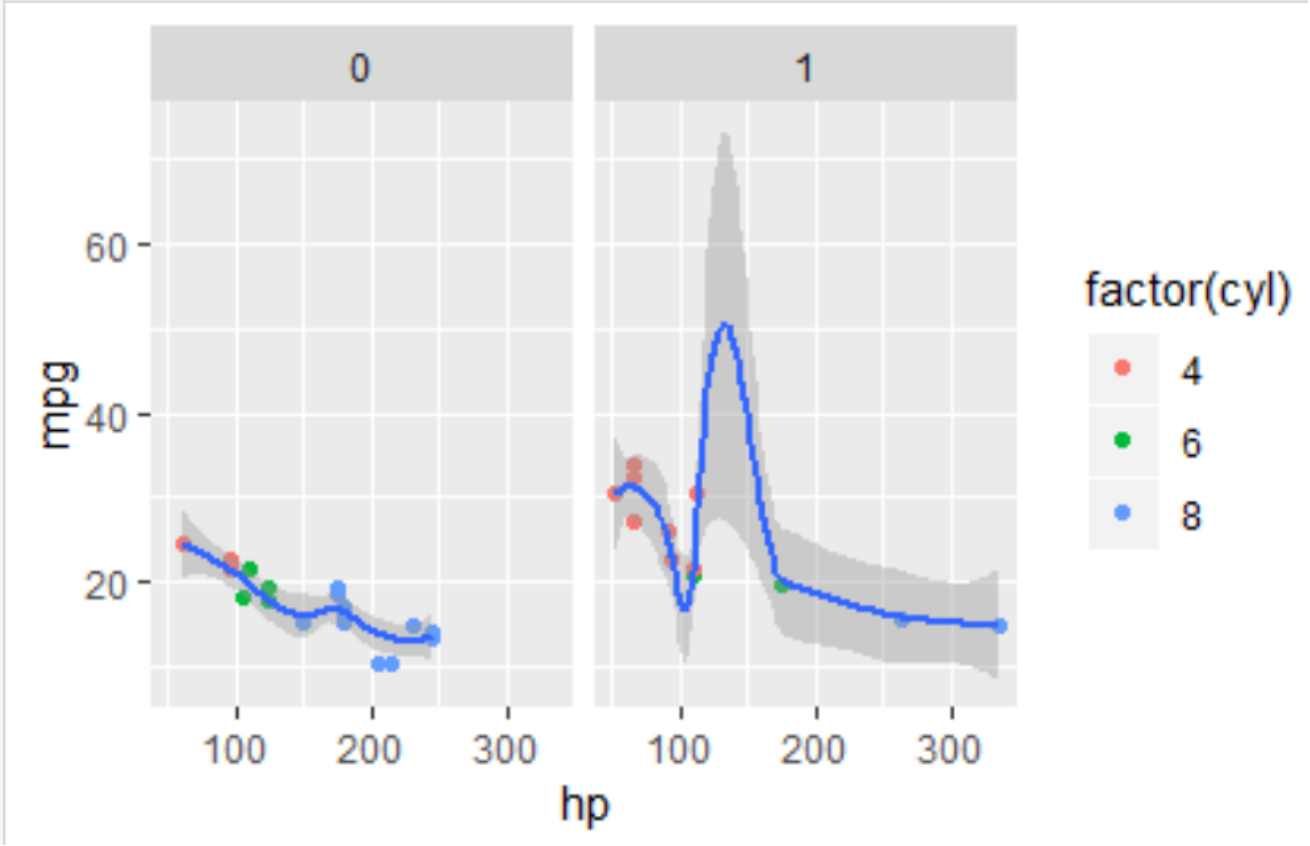
Facet

```
ggplot(df1,aes(hp))+
  geom_dotplot(aes(color=factor(cyl)))+
  facet_grid(am~vs,margins = TRUE)
```



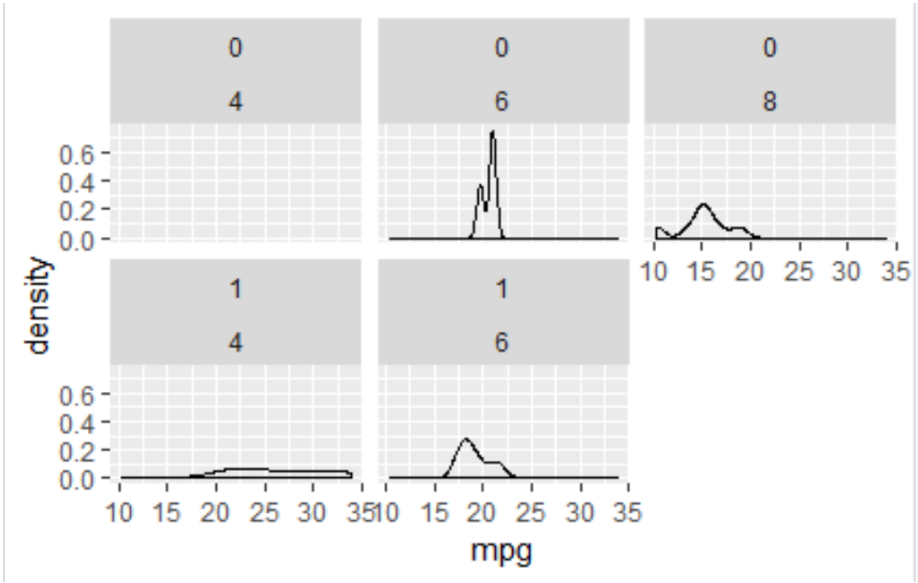
Facet

```
ggplot(df1,aes(hp,mpg))+
  geom_point(aes(color=factor(cyl)))+
  facet_grid(.~am)+
  geom_smooth()
```

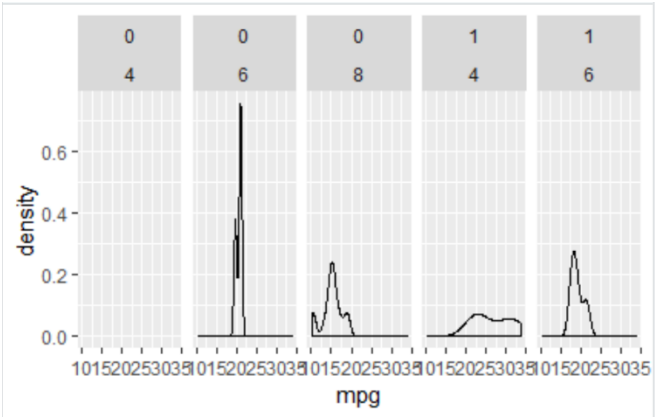


Facet

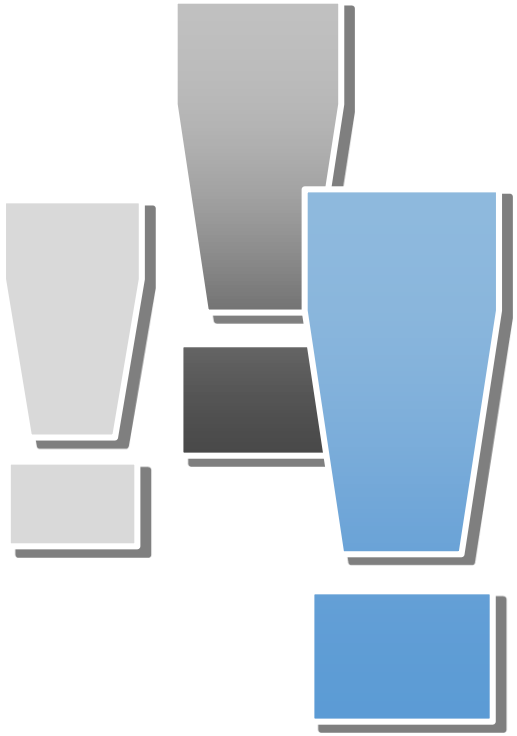
```
ggplot(df1,aes(mpg))+
  geom_density()+
  facet_wrap(~factor(vs)+factor(cyl))
```




```
ggplot(df1,aes(mpg))+
  geom_density()+
  facet_wrap(~factor(vs)+factor(cyl),nrow = 1)
```



Спасибо за внимание!



Шевцов Василий Викторович

shevtsov_vv@rudn.university
+7(903)144-53-57