

## Лабораторная работа №1. Построение графиков

Лабораторная работа служит для получения практических навыков по изучению Модуля 1 данного курса.

**Цель работы:** в зависимости от поставленной цели исследования подобрать тип визуализации и построить графики.

Номер варианта соответствует номеру в журнале преподавателя. Если номер больше 10, например 11 или 21 – выбираете вариант номер 1 и т.д.

Все наборы данных, используемые в работе, доступны в разделе лабораторной работы № 1.

**Ход работы:**

### ЧАСТЬ 1

diamonds – набор данных, содержащий сведения о 53940 бриллиантах. Используется 10 переменных: price, carat, cut, color, clarity, x, y, z, depth и table (рисунок 1).

Описание данных можно посмотреть, перейдя по ссылке:

<https://www.kaggle.com/shivam2503/diamonds>

	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
...	...	...	...	...	...	...	...	...	...	...
53936	0.72	Ideal	D	SI1	60.8	57.0	2757	5.75	5.76	3.50
53937	0.72	Good	D	SI1	63.1	55.0	2757	5.69	5.75	3.61
53938	0.70	Very Good	D	SI1	62.8	60.0	2757	5.66	5.68	3.56
53939	0.86	Premium	H	SI2	61.0	58.0	2757	6.15	6.12	3.74
53940	0.75	Ideal	D	SI2	62.2	55.0	2757	5.83	5.87	3.64

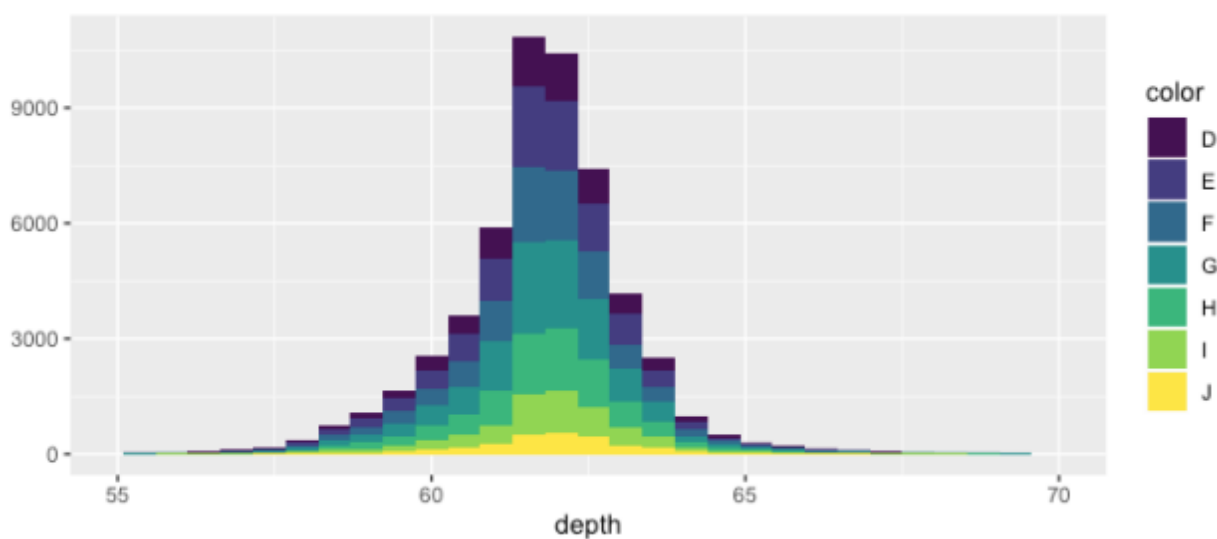
53940 rows × 10 columns

Рисунок 1 – Структура набора данных diamonds

**Задание 1.** Напишите программный код, строящий следующие графики по набору данных **diamonds** (допустимо, графики могут незначительно отличаться от приведенных здесь).  
 Дайте краткое описание, что показывает график.

Вариант	1	2	3	4	5	6	7	8	9	10
№ графиков	1,5	2,6	3,7	4,8	5,9	6,10	7,4	8,2	9,3	10,1

**График № 1**



**График № 2**

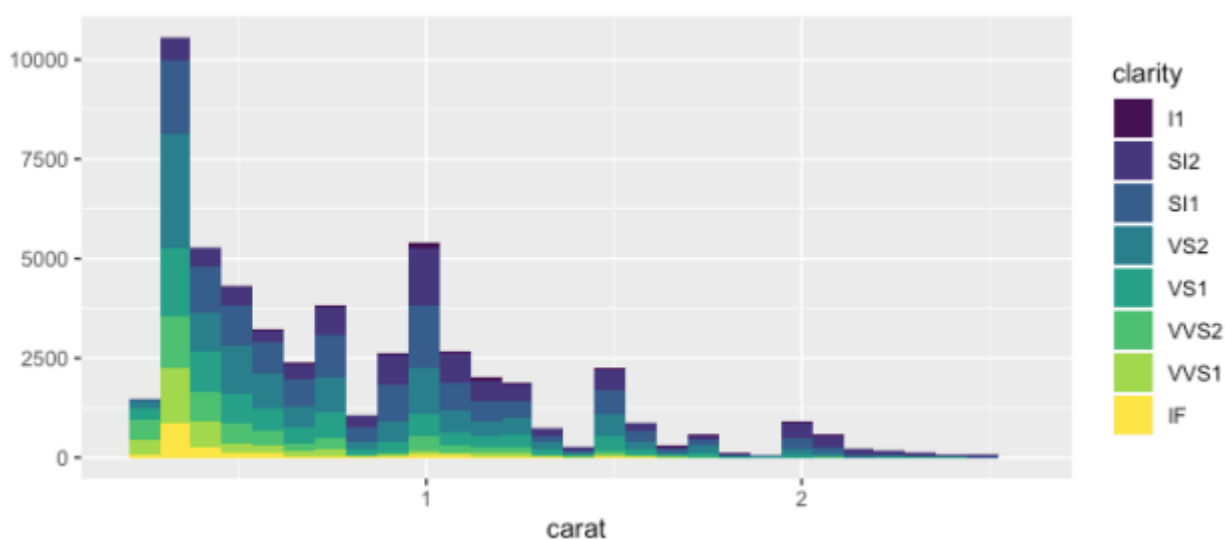


График № 3

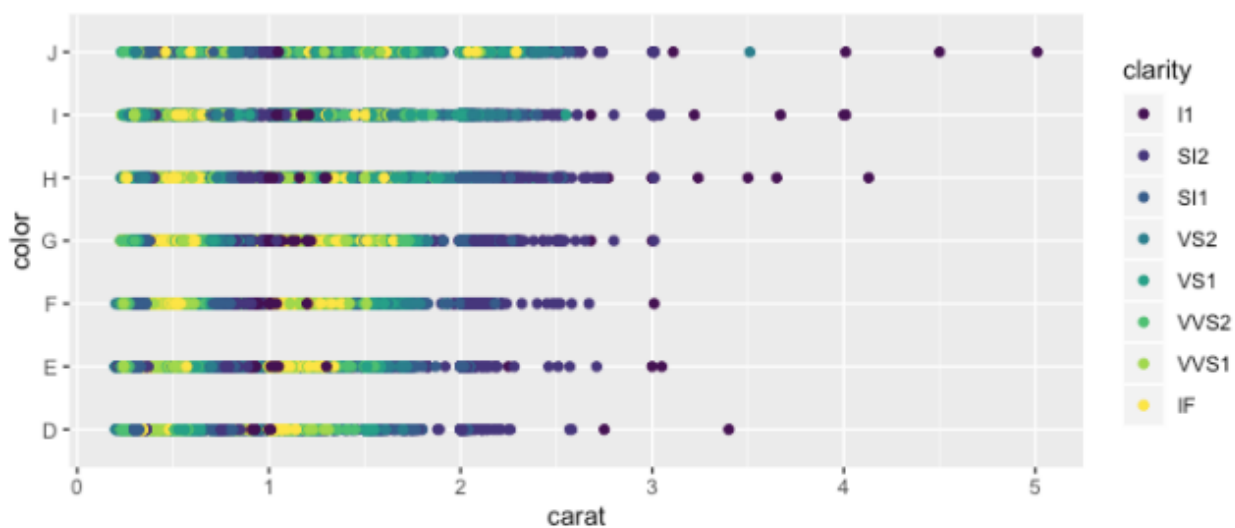


График № 4

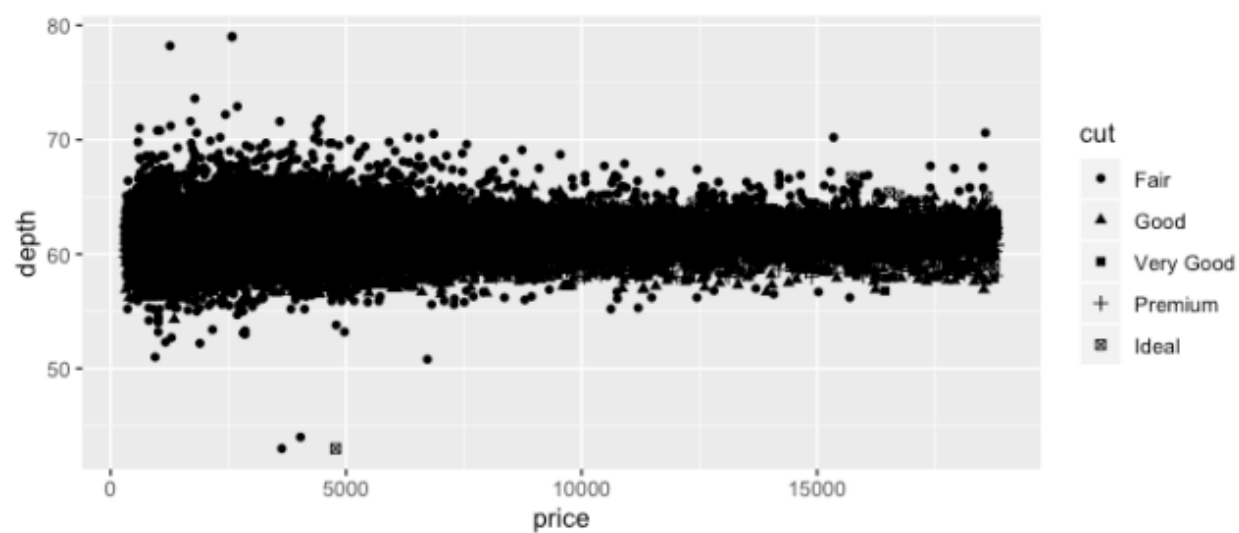


График № 5

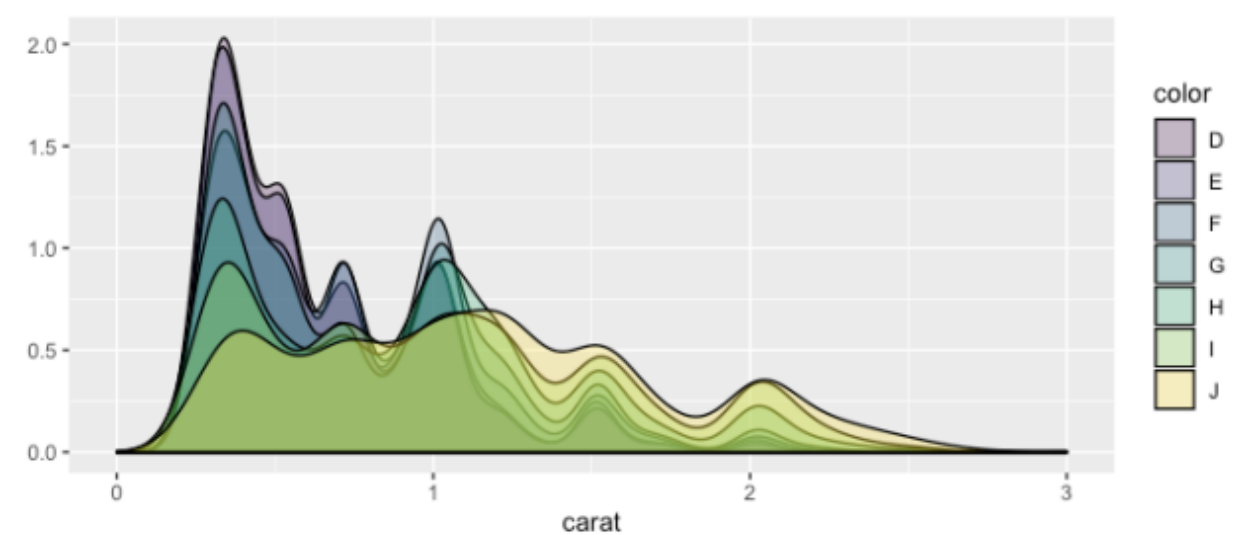


График № 6

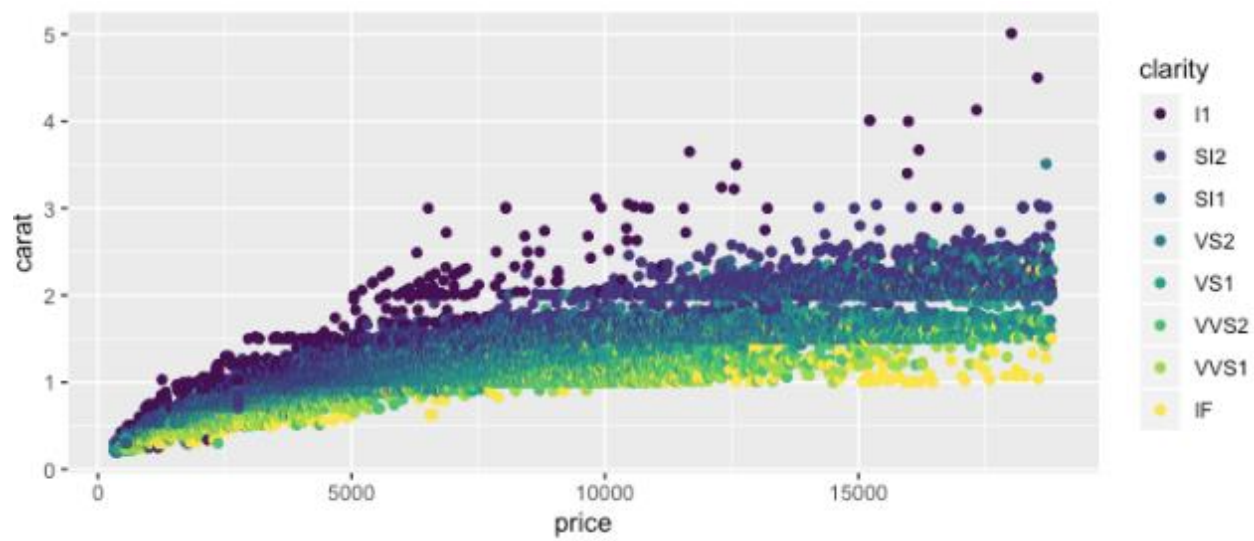


График № 7

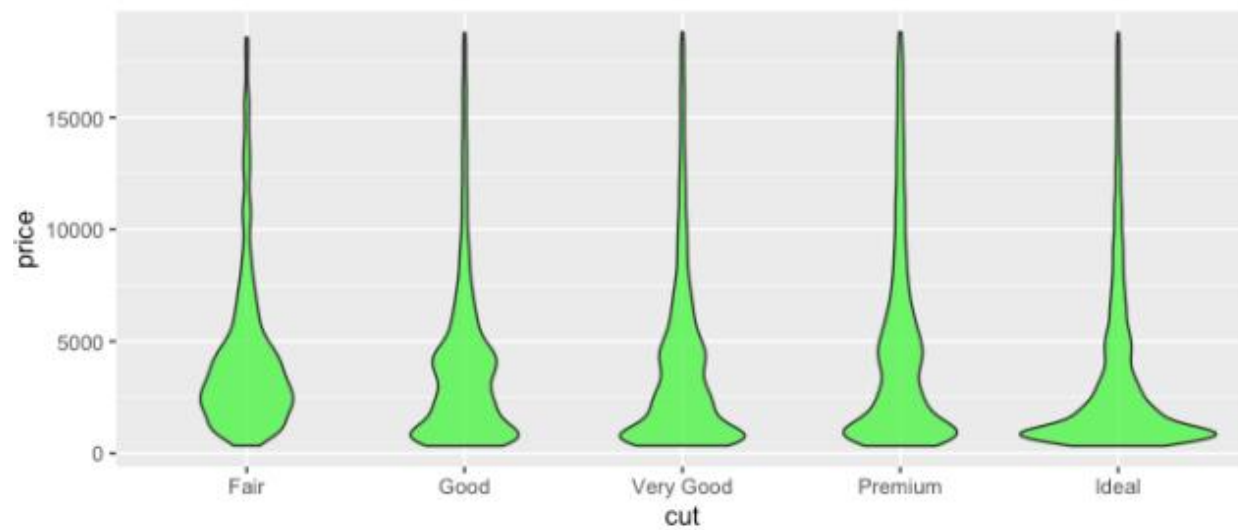


График № 8

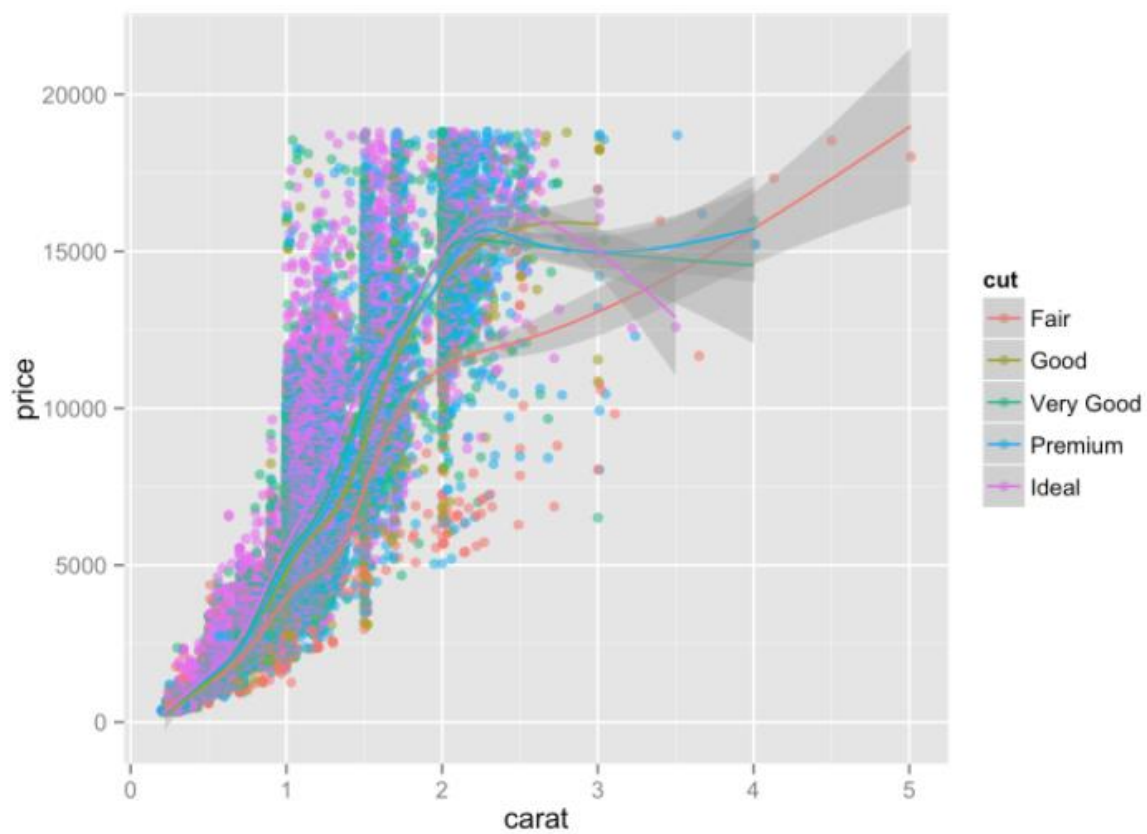


График № 9

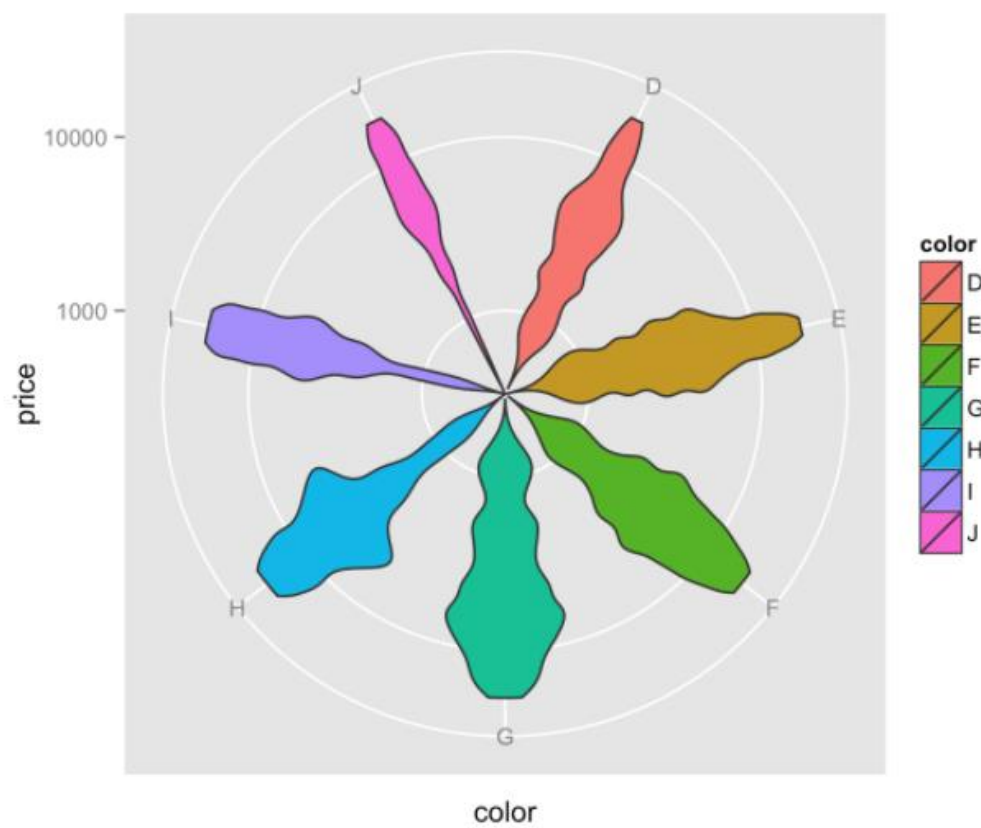
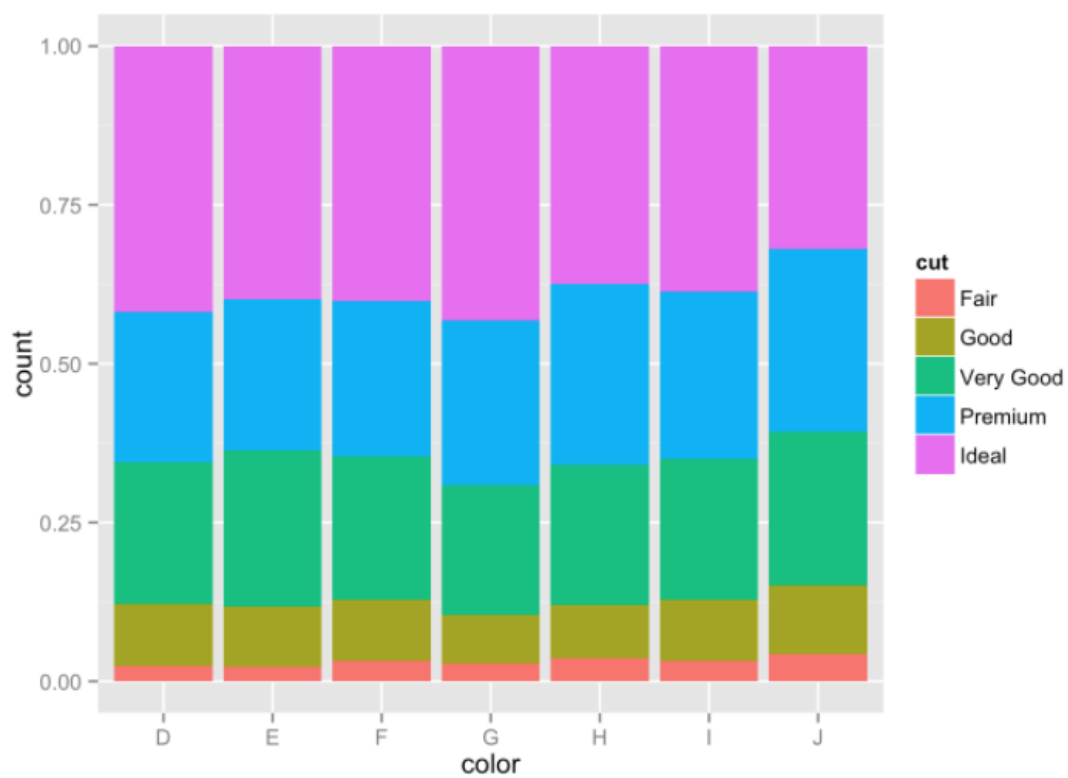


График № 10



## ЧАСТЬ 2

mtcars – набор данных, содержащий данные о характеристиках 32 автомобилей из журнала Motor Trend за 1974. Используется 12 переменных: model, mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear и carb (рисунок 2).

Описание данных можно посмотреть, перейдя по ссылке:

<https://www.kaggle.com/vik2012kvs/mt-cars>

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
6	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
7	Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
8	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
9	Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
10	Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4

Рисунок 2 – Структура набора данных mtcars

**Задание 2.** Напишите программный код, строящий следующие графики по набору данных **mtcars** (допустимо, графики могут незначительно отличаться от приведенных здесь). Дайте краткое описание, что показывает график.

Вариант	1	2	3	4	5	6	7	8	9	10
№ графиков	1,5	2,6	3,7	4,8	5,9	6,10	7,4	8,2	9,3	10,1

График № 1

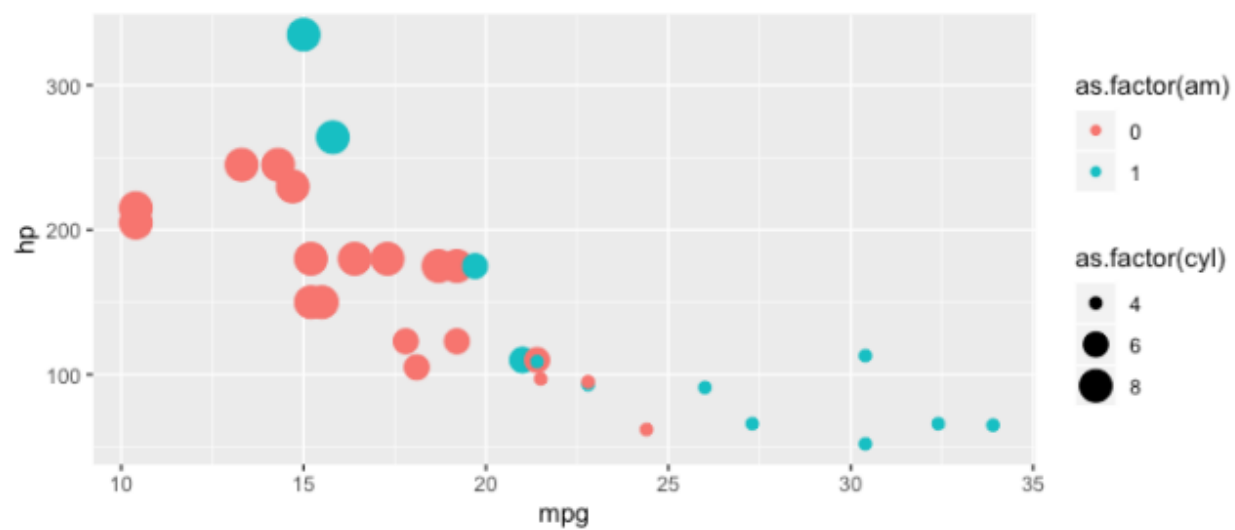


График № 2

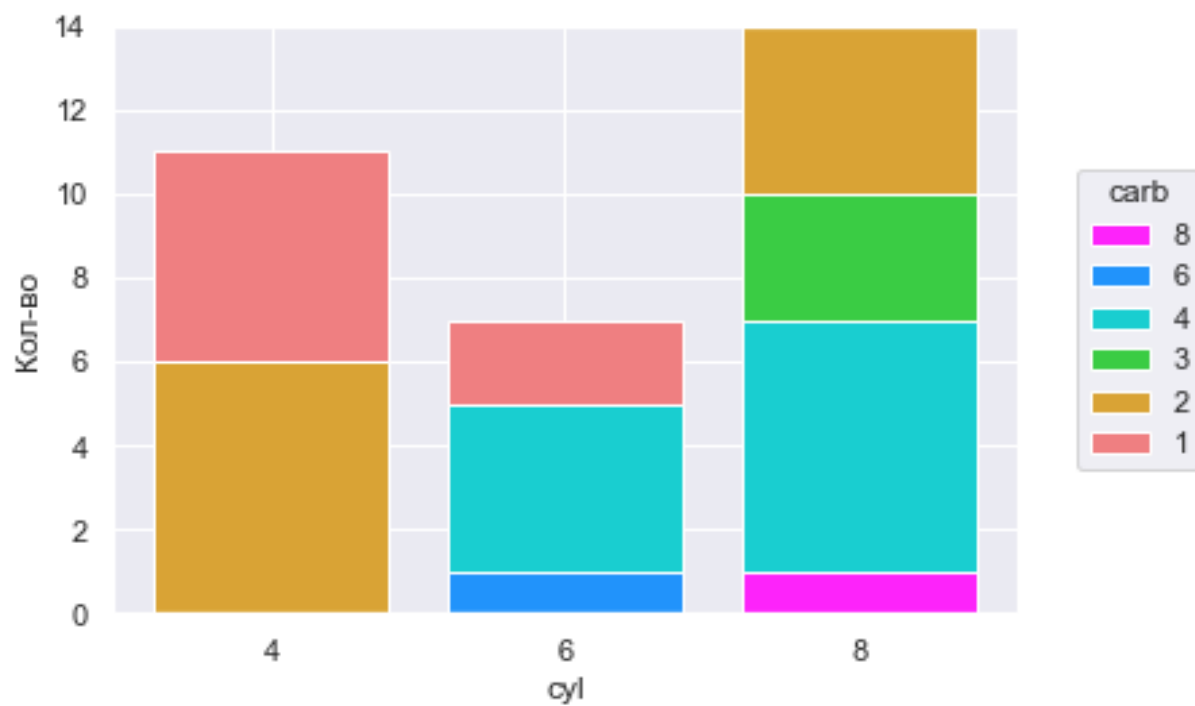




График № 3

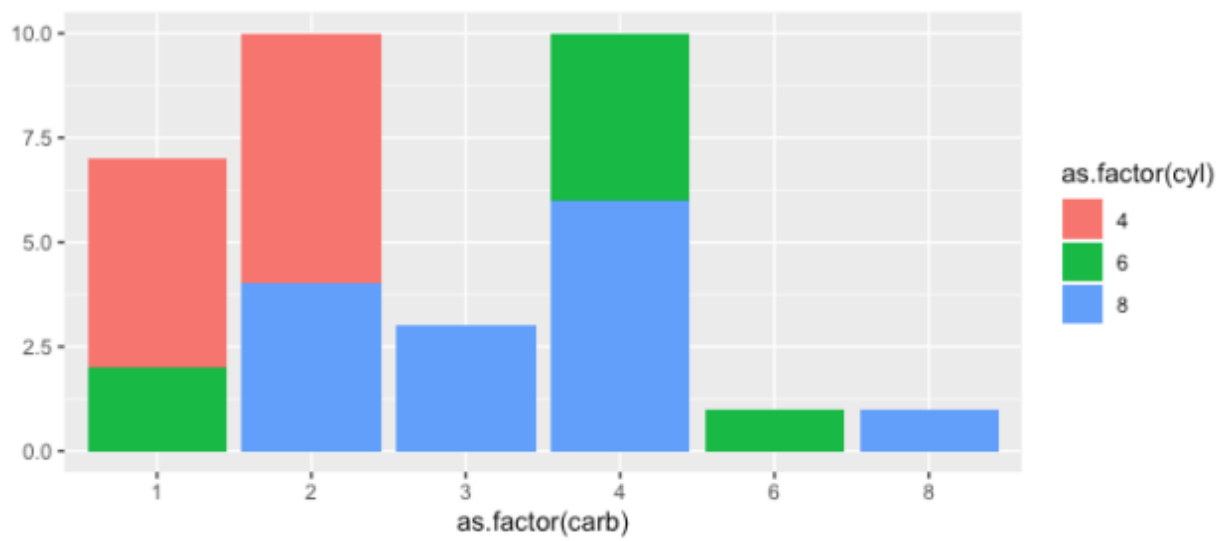


График № 4

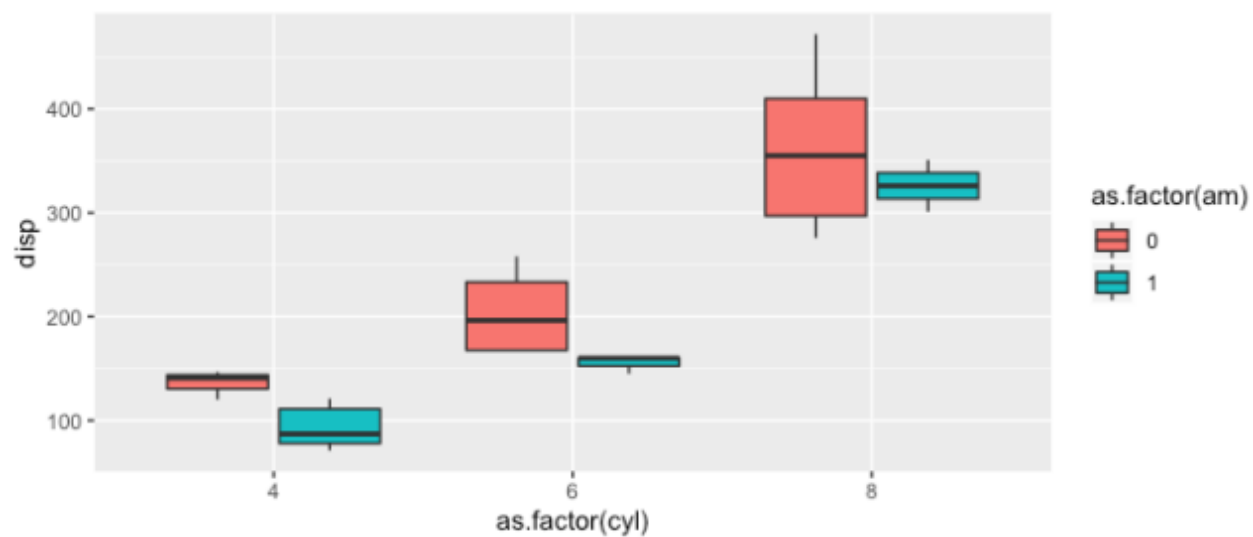


График № 5

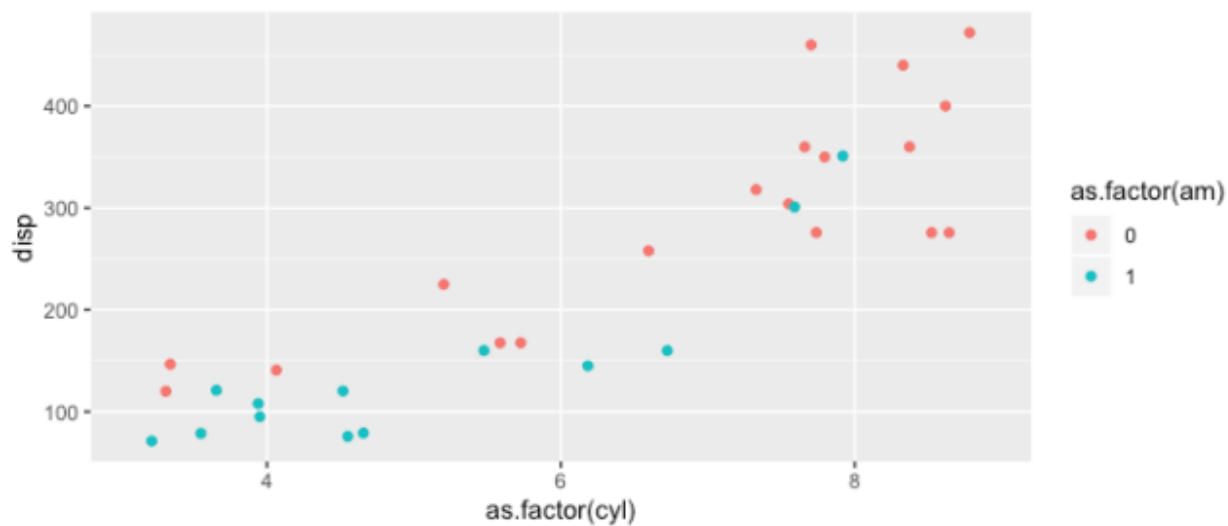


График № 6

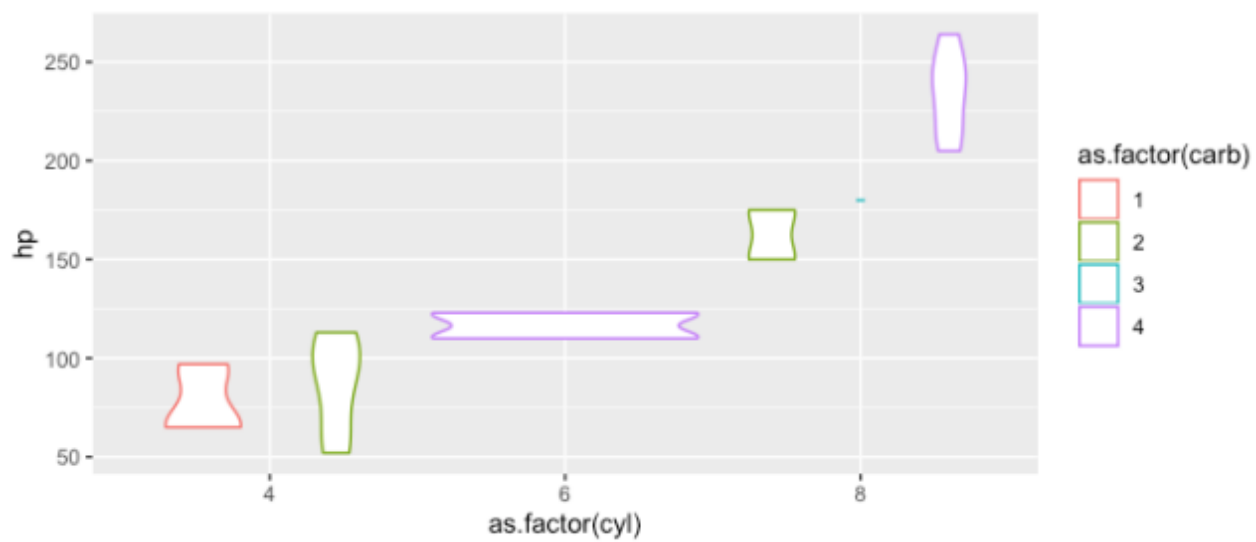


График № 7

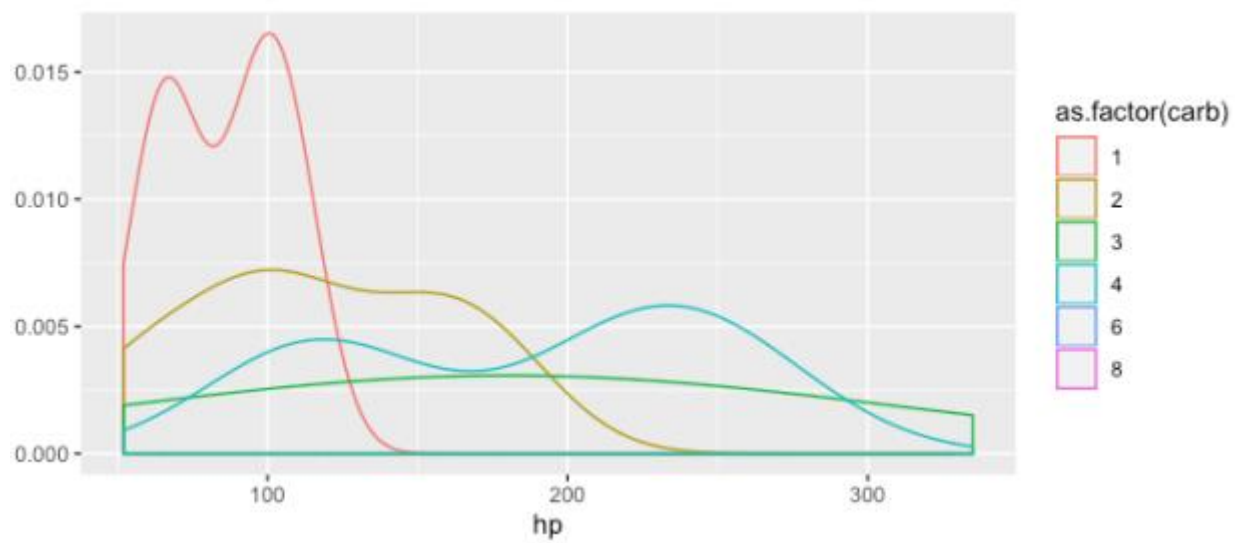


График № 8

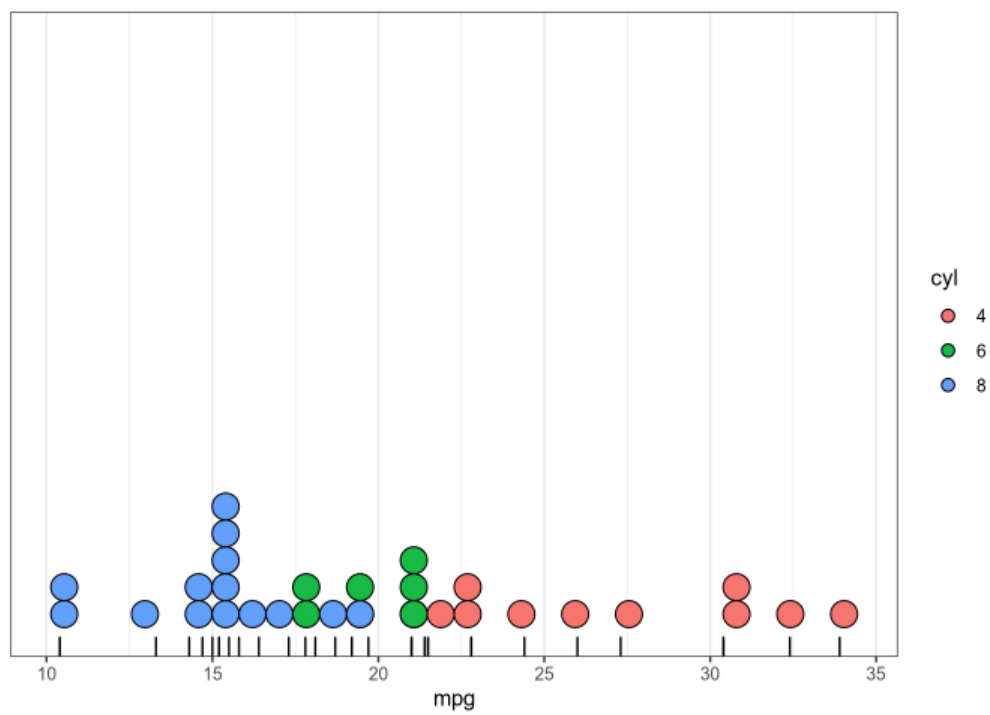


График № 9

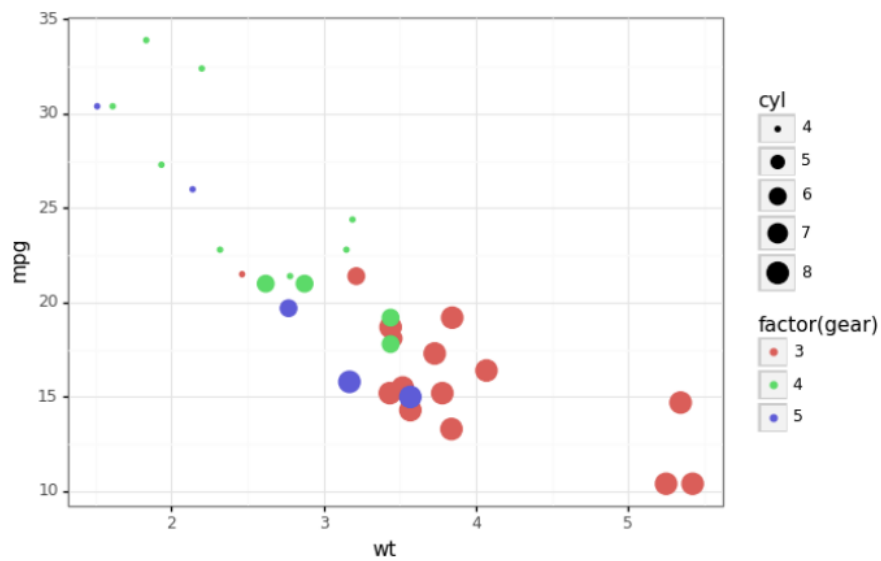
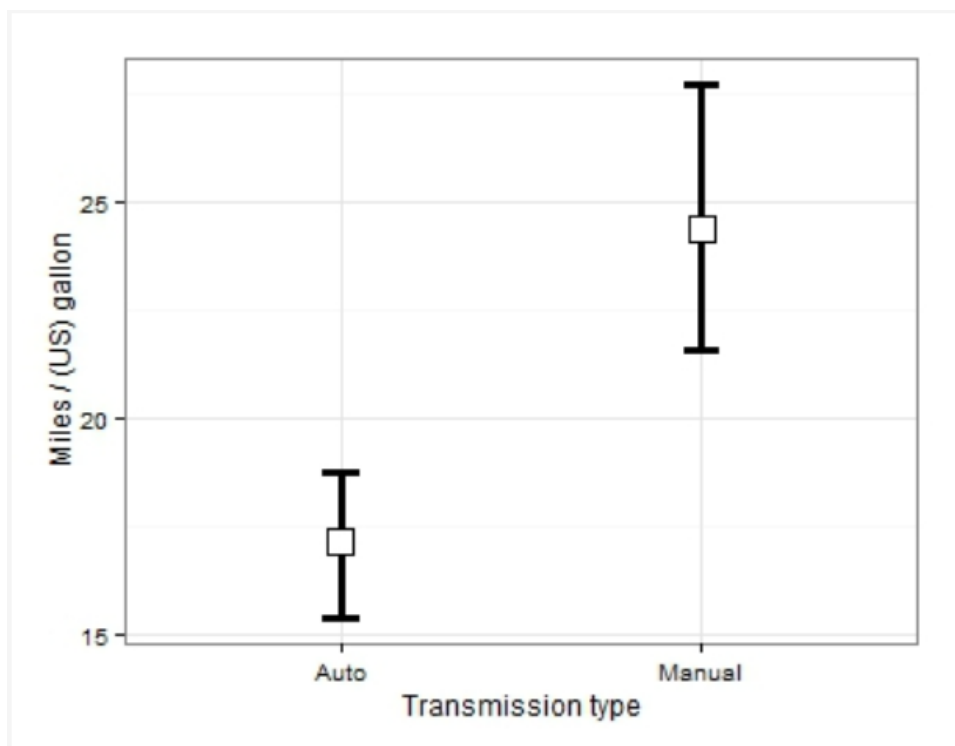


График № 10



Transmission type (am), miles\_per\_gallon (mpg).

### ЧАСТЬ 3

Постройте два произвольных графика, показав умение использовать различные настройки для набора данных по заданному варианту. Дайте описание набору данных и то, что показывает изображенный график.

Описание данных можно посмотреть, перейдя по ссылке:

<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

Вариант	Набор данных
1	edcCO2
2	ChickWeight.
3	DNase
4	LifeCycleSavings
5	Loblolly
6	Orange
7	OrchardSprays
8	Puromycin
9	Seatbelts
10	Theoph
11	ToothGrowth
12	USArrests
13	freeny
14	infert
15	iris
16	longley
17	quakes
18	rock
19	stackloss
20	swiss
21	trees
22	economics
23	faithfuld
24	midwest
25	txhousing

#### Задание:

- 1) Загрузите данные согласно номеру Вашего варианта.
- 2) Подготовьте краткое описание – о чем Ваши данные, какова структура данных.
- 3) Постройте:
  - а) 5 нижеприведенных типов диаграммы, в основе которых лежат разные типы сравнений.
  - б) 2 типа – на свой выбор.Настройте оформление диаграмм (заголовков, цвета, маркеры, подписи).
- 4) Напишите вывод, в котором отразите найденные закономерности.

**Программный код для представленных примеров в файле  
Lab1VizualzationCoreAI.ipynb:**

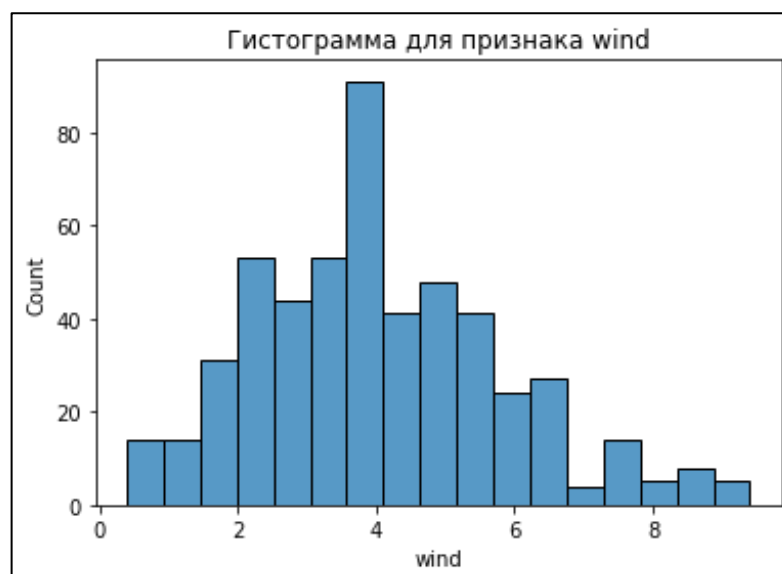
1. Постройте круговую диаграмму для подходящего признака из вашего набора данных. (можно, например, использовать метод `matplotlib.pyplot.pie`). Настройте диаграмму так, чтобы она была похожа на приведенный пример.



Пример

```
#Круговая диаграмма
plt.figure(figsize=(6,6))
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)
plt.title('Распределение пожаров по дням недели')
plt.show()
```

2. Постройте гистограмму распределения для подходящего признака из вашего набора данных. (можно, например, использовать метод `seaborn.histplot`). Настройте диаграмму так, чтобы она была похожа на приведенный пример.



### Пример

```
#Гистограмма
sns.histplot(df, x="wind")
plt.title("Гистограмма для признака wind")
plt.show()
```

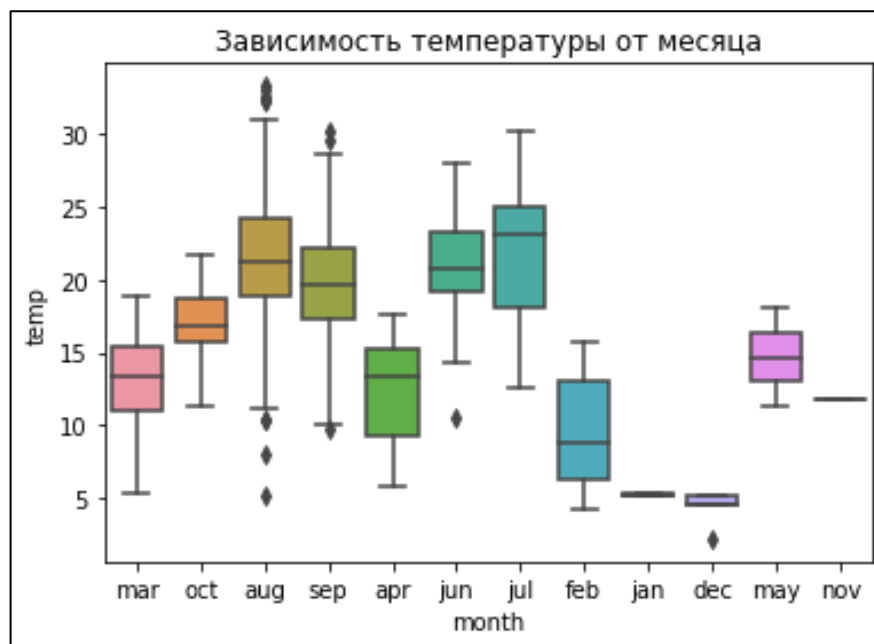
3. Постройте точечную диаграмму для подходящих признаков из вашего набора данных. (можно, например, использовать метод `seaborn.scatterplot`). Настройте диаграмму так, чтобы она была похожа на приведенный пример.



### Пример

```
#Точечная диаграмма
sns.scatterplot(data=df, x="RH", y="temp", color='orange')
plt.title('Точечная диаграмма по признакам RH и temp')
plt.show()
```

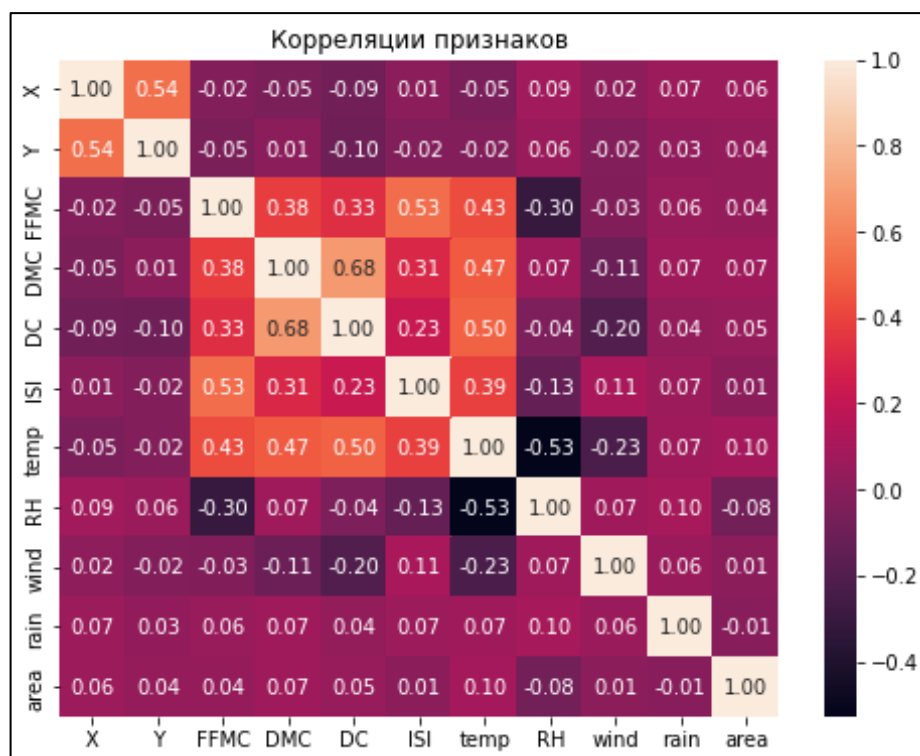
4. Постройте диаграмму «ящик с усами» (boxplot) для подходящего признака из вашего набора данных. (можно, например, использовать метод `seaborn.boxplot`). Настройте диаграмму так, чтобы она была похожа на приведенный пример.



Пример

```
#Ящик с усами
sns.boxplot(data=df, x="month", y="temp")
plt.title("Зависимость температуры от месяца")
plt.show()
```

5. Постройте тепловую карту корреляций для всех числовых признаков из вашего набора данных. (можно, например, использовать методы `seaborn.heatmap` и `df.corr`). Настройте диаграмму так, чтобы она была похожа на приведенный пример.





## Пример

```
#Данные
data = df.corr()

#Тепловая карта
plt.figure(figsize=(8,6))
sns.heatmap(data,annot=True, fmt=".2f")
plt.title('Корреляции признаков')
plt.show()
```

## Подготовка отчета

Допустима подготовка отчета в формате *Jupyter Notebook* (или аналогичной среде) – с подробными комментариями к коду и результатам визуализации. Комментарии должны содержать не только факт выполнения – построена такая-то диаграмма, но также что Вы хотели показать, используя ту или иную диаграмму, и что получили в итоге.

В качестве результата – на проверку преподавателю – загружается файл-отчет с исходным кодом (например, \*.ipynb), а также отчет в формате html. Загружаются также все массивы данных, на основе которых была проведена работа.

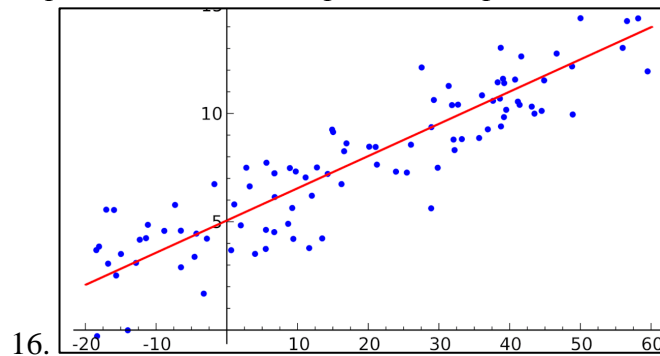
## Перечень примерных вопросов на защите работы:

1. Какие базовые диаграммы доступны в библиотеке *matplotlib*?
2. Какие существуют еще библиотеки визуализации на языках *R* и *Python*?
3. На какой библиотеке основана библиотека *Seaborn*?
4. Какой метод из библиотеки *matplotlib* позволяет быстро взглянуть на данные с помощью построения двумерных точечных диаграмм и гистограмм распределения?
5. Что такое ус, медиана и выброс в диаграмме «ящик с усами»?
6. Зачем на гистограмму распределения добавляют функцию плотности распределения?
7. Какие выводы можно сделать по тепловой карте корреляций признаков?
8. Какой тип диаграммы лучше подходит для визуализации зависимости категориального признака от вещественного (точечная диаграмма или «ящик с усами») и почему?
9. Назовите 5 основных типов диаграмм.
10. Какие типы диаграмм подходят для корреляционного типа сравнения?
11. Какому типу сравнения соответствует круговая диаграмма?
12. В каких случаях используется линейчатая диаграмма, а в каких – столбчатая?
13. В каком случае для временного сравнения рекомендуется использовать линейчатую диаграмму, а в каких график?
14. Определите тип сравнения и подходящую диаграмму для предложенных данных.

Дата	Кол-во посетителей
26.11.21	145
27.11.21	178
28.11.21	354

29.11.21	240
30.11.21	198
01.12.21	230

15. Определите тип сравнения и тип диаграммы по предложенной диаграмме.



17. По описанию гипотезы, лежащей в основе диаграммы, определите тип сравнения и подходящий тип диаграммы. Описание: «Партия А одержала уверенную победу на выборах, получив 55% голосов».