

Визуализация данных реальной задачи

Иван Сомов

МГУ им. М.В. Ломоносова

Факультет вычислительной математики и кибернетики

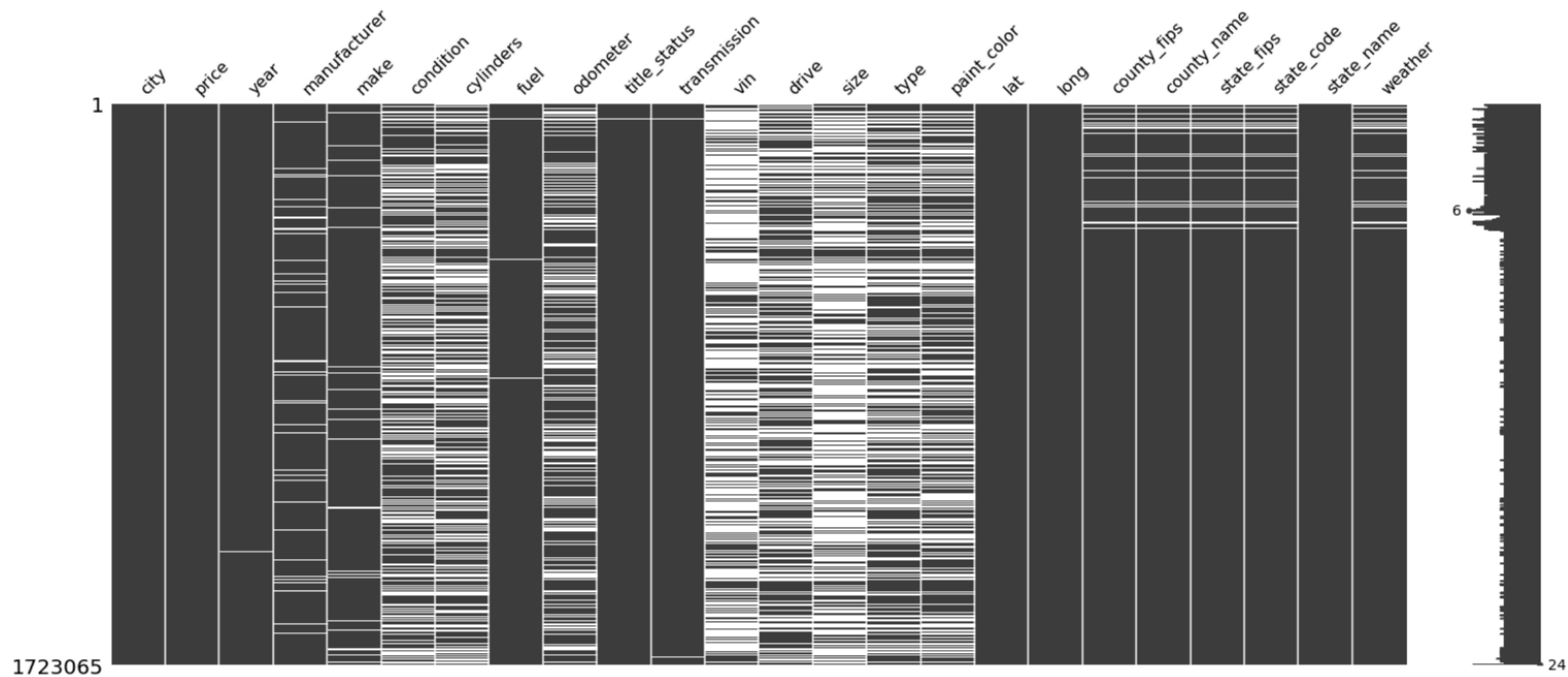
Данные

- Американская онлайн доска объявлений Craigslist (типа “из рук в руки”, “молоток.ру”, “авито” и тд)
- Взяты данные из объявлений о продаже машин

	city	price	year	manufacturer	make	condition	cylinders	fuel	odometer	title_status	transmission	vin	drive
3	marshall	2800	2004.0	ford	taurus	good	6 cylinders	gas	168591.0	clean	automatic	1FAFP55UO4G113464	fwd
140	reading	4480	2010.0	honda	insight	good	4 cylinders	hybrid	110000.0	clean	automatic	JHMZE2H76AS003740	fwd
157	chambersburg	4500	2005.0	chevy	2500 hd	good	8 cylinders	gas	204000.0	clean	automatic	Xxxxxxxxxxx	rwd
203	rapidcity	38500	2017.0	chevrolet	silverado	like new	8 cylinders	gas	44710.0	clean	automatic	3GCUKREC8HG363009	4wd
243	salina	800	2001.0	ford	escort zx2	fair	4 cylinders	gas	93400.0	clean	automatic	3FAFP11371R241940	fwd

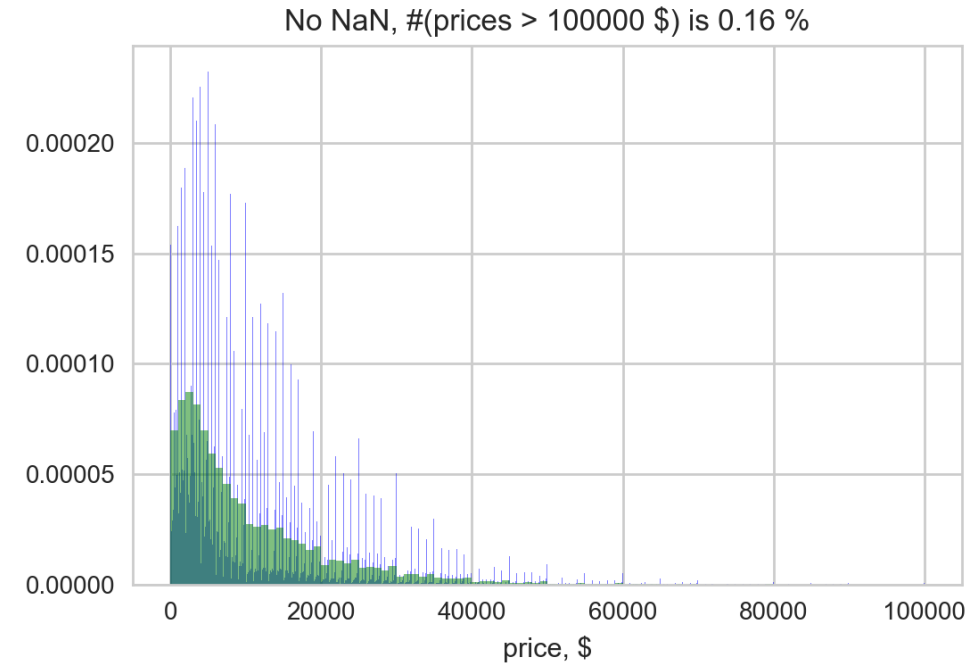
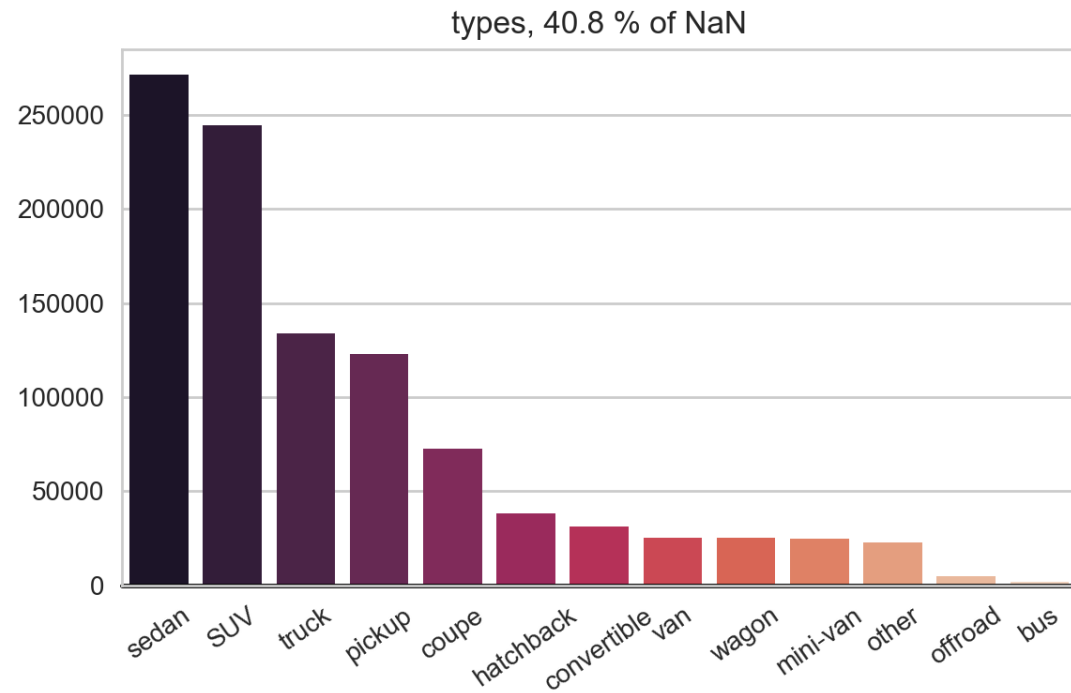
type	paint_color	state_name	weather
sedan	grey	iowa	47.0
hatchback	silver	Pennsylvania	53.0
truck	grey	Pennsylvania	53.0
offroad	red	South Dakota	38.0
coupe	brown	Kansas	51.0

Данные

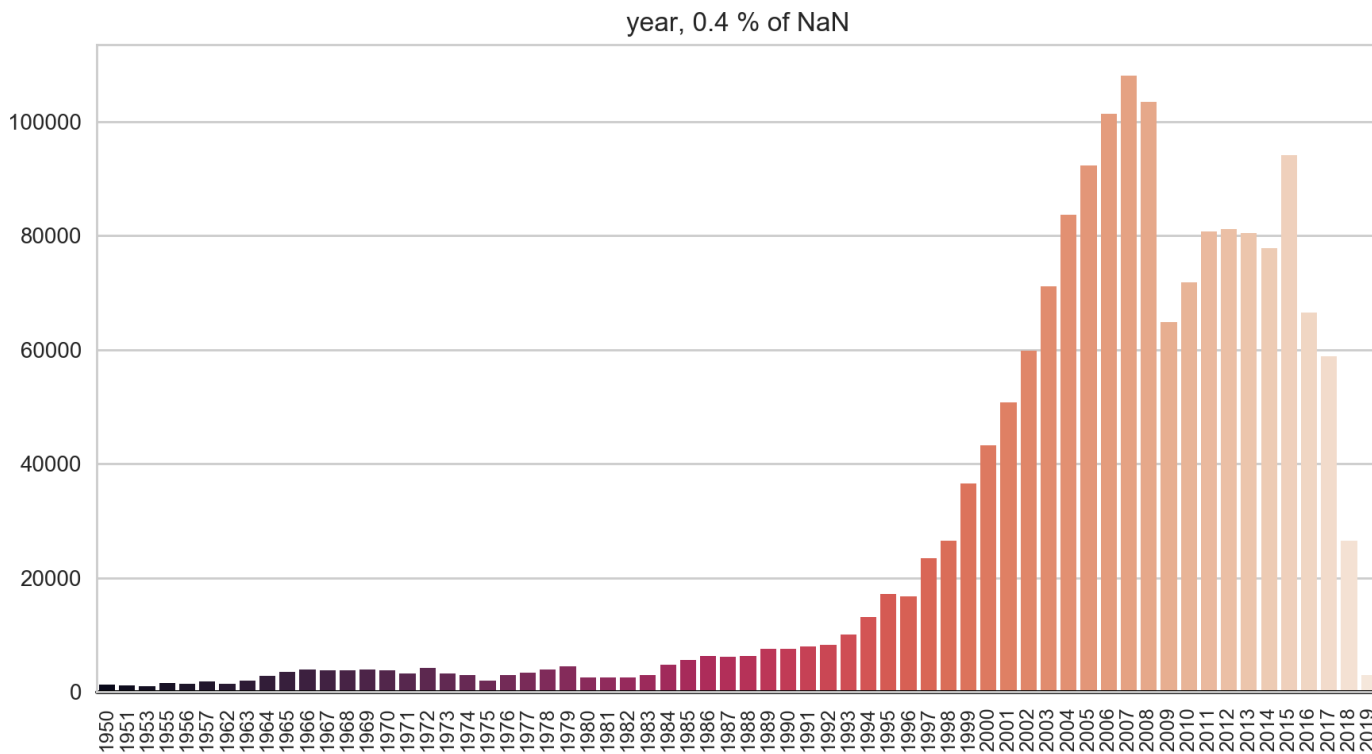


- pip install missingno

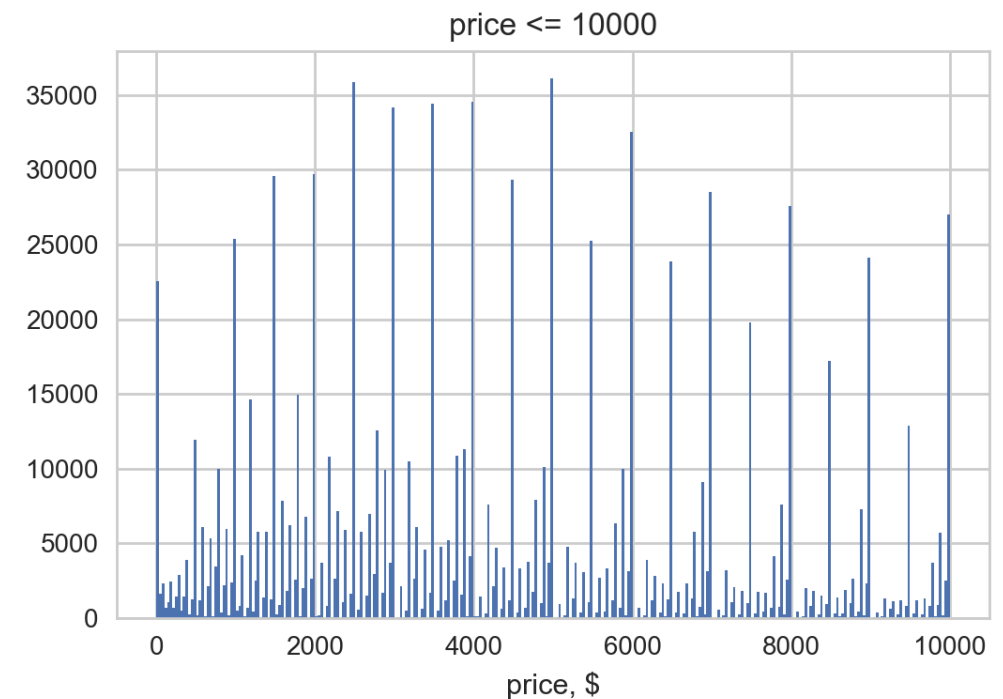
Посмотрим на базовые графики



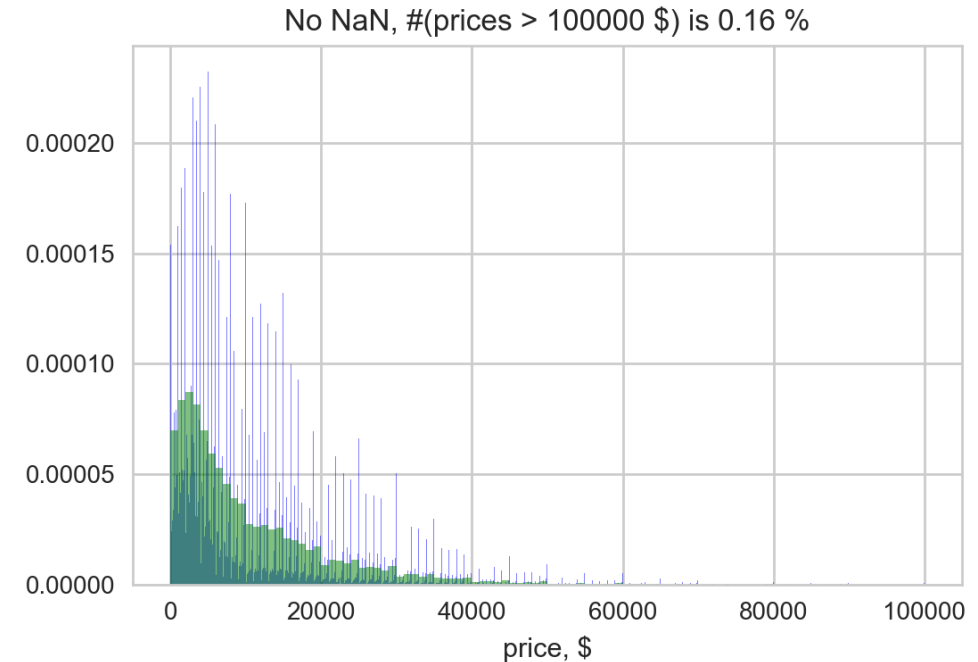
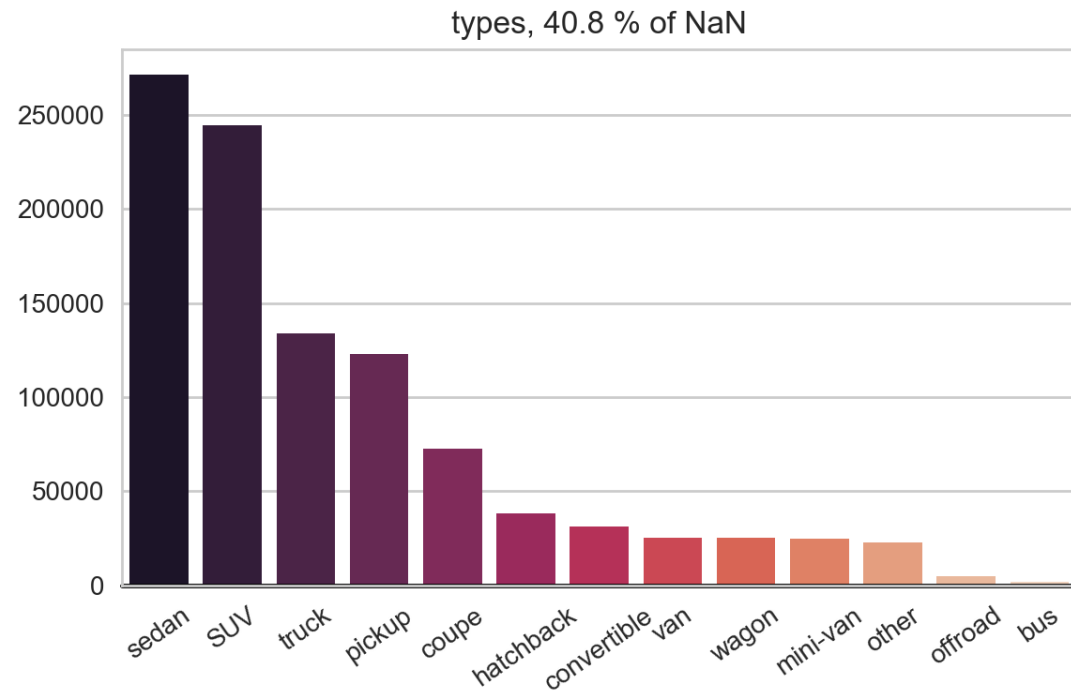
Минимальная цена 1\$. Что за высокие столбики?



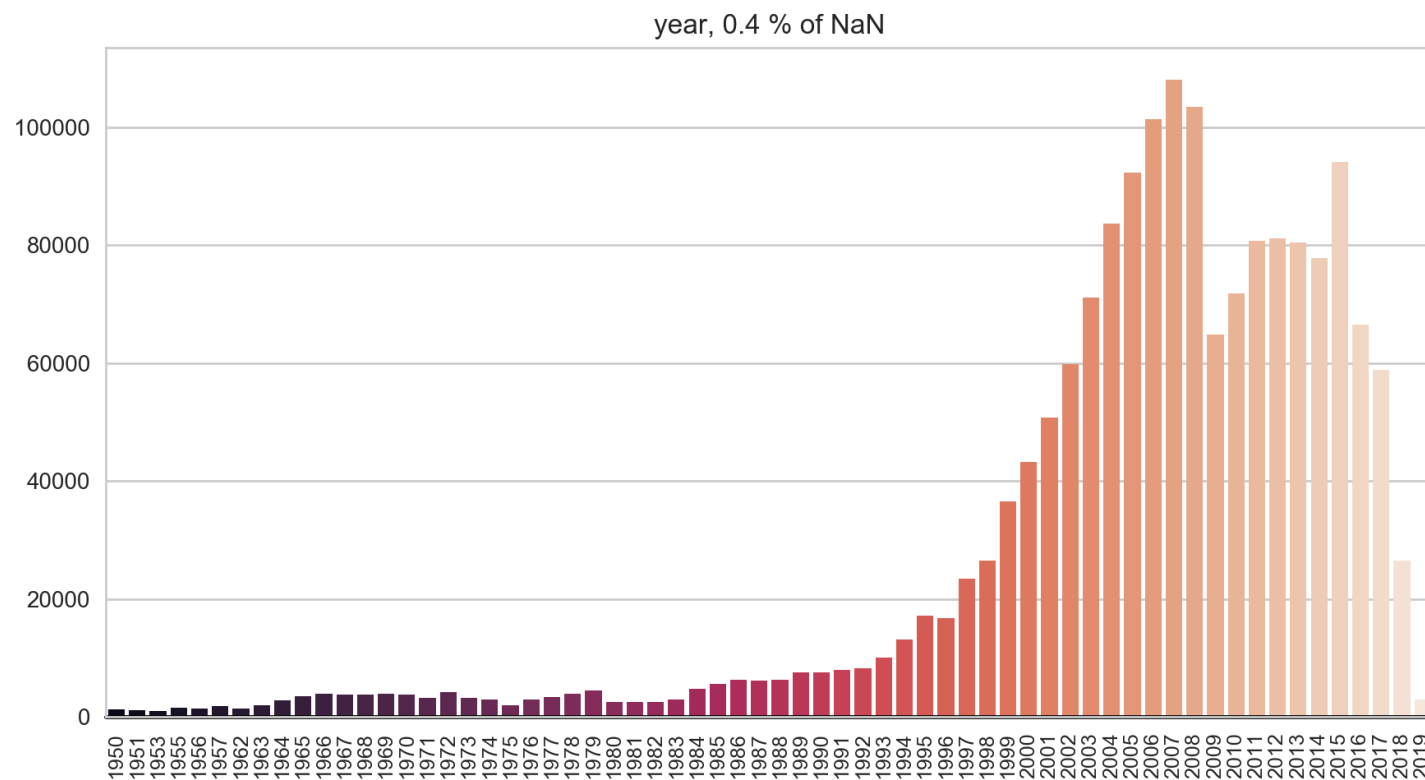
Что за обвал в нулевых?



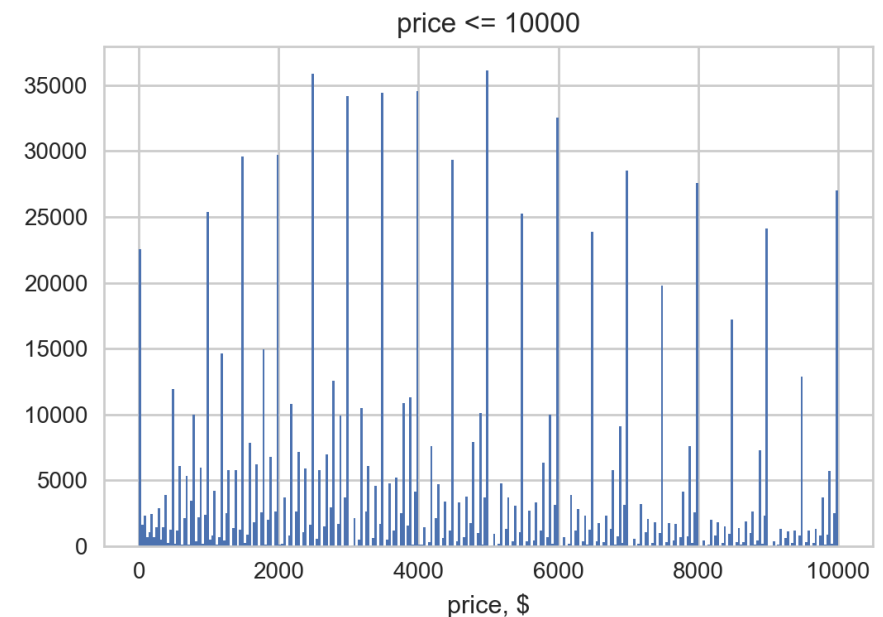
Посмотрим на базовые графики



Минимальная цена 1\$. Что за высокие столбики?
Круглые цифры

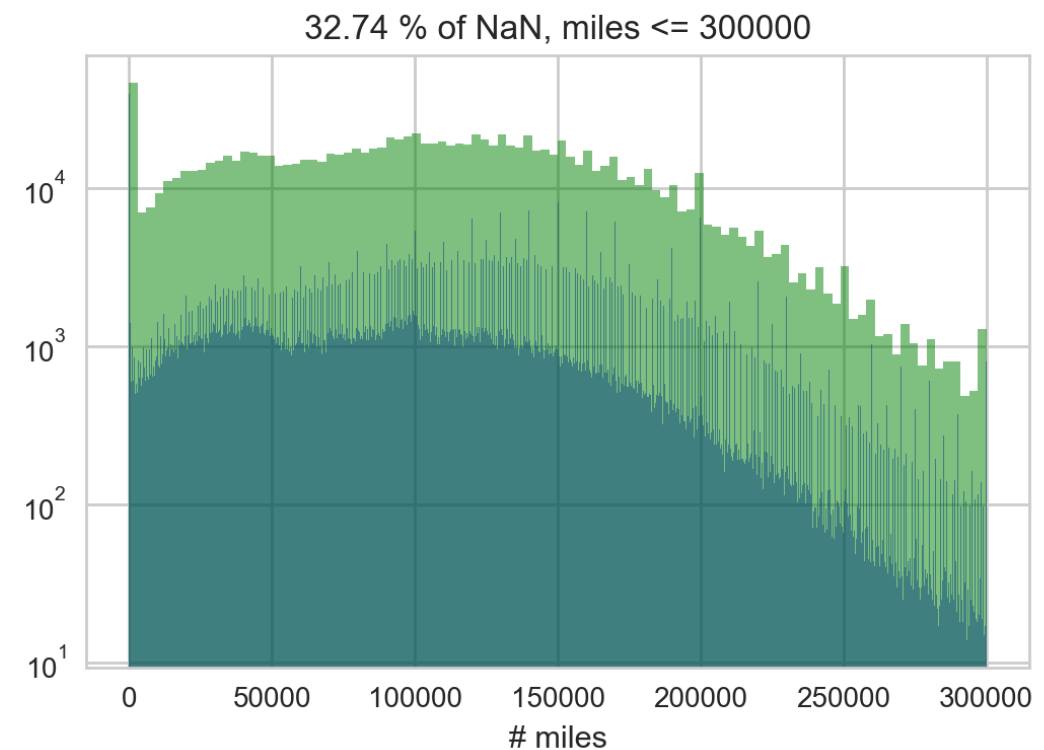
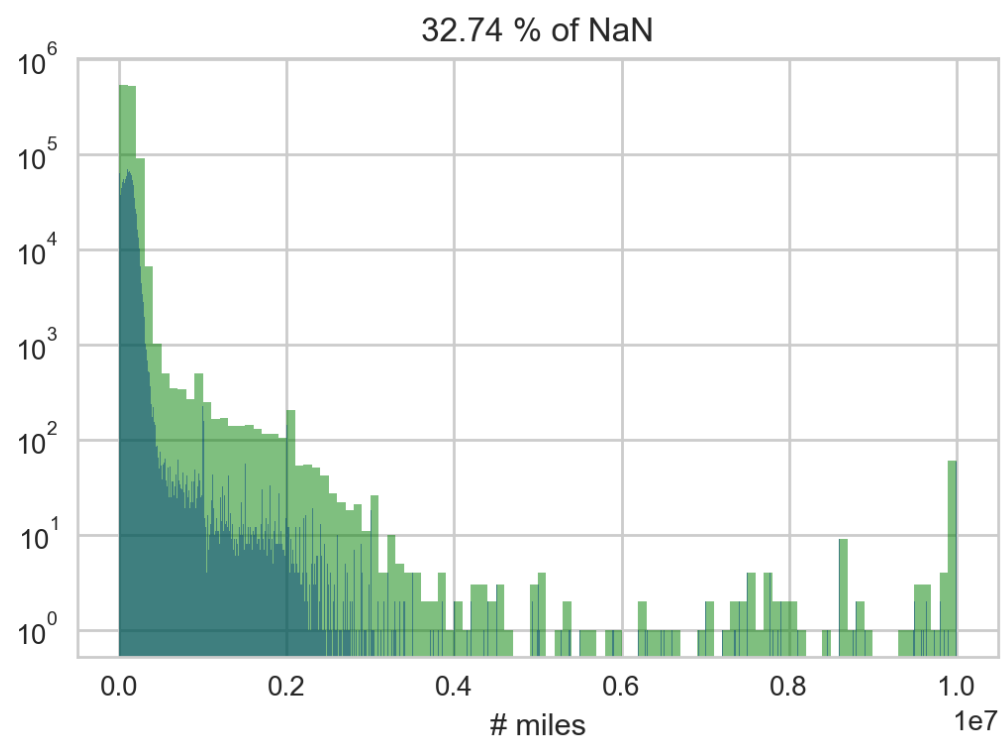


Что за обвал в нулевых? Кризис 2008
С 302 г. до 1949 г. слишком редкие объявления



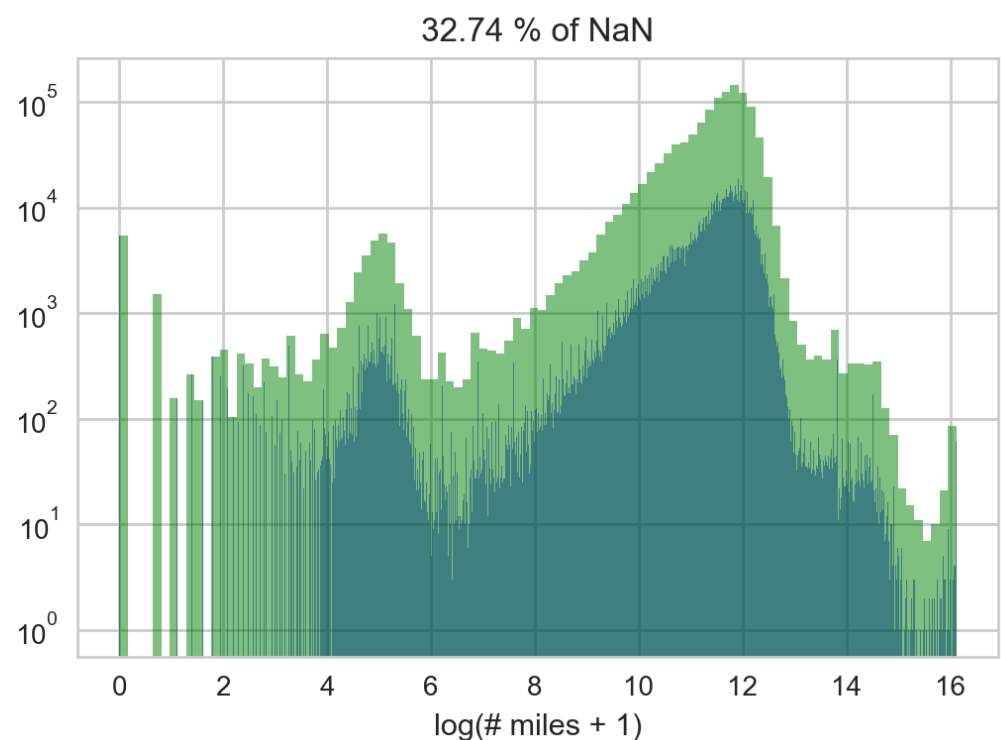
Упоминаем NaN!!

Посмотрим на базовые графики



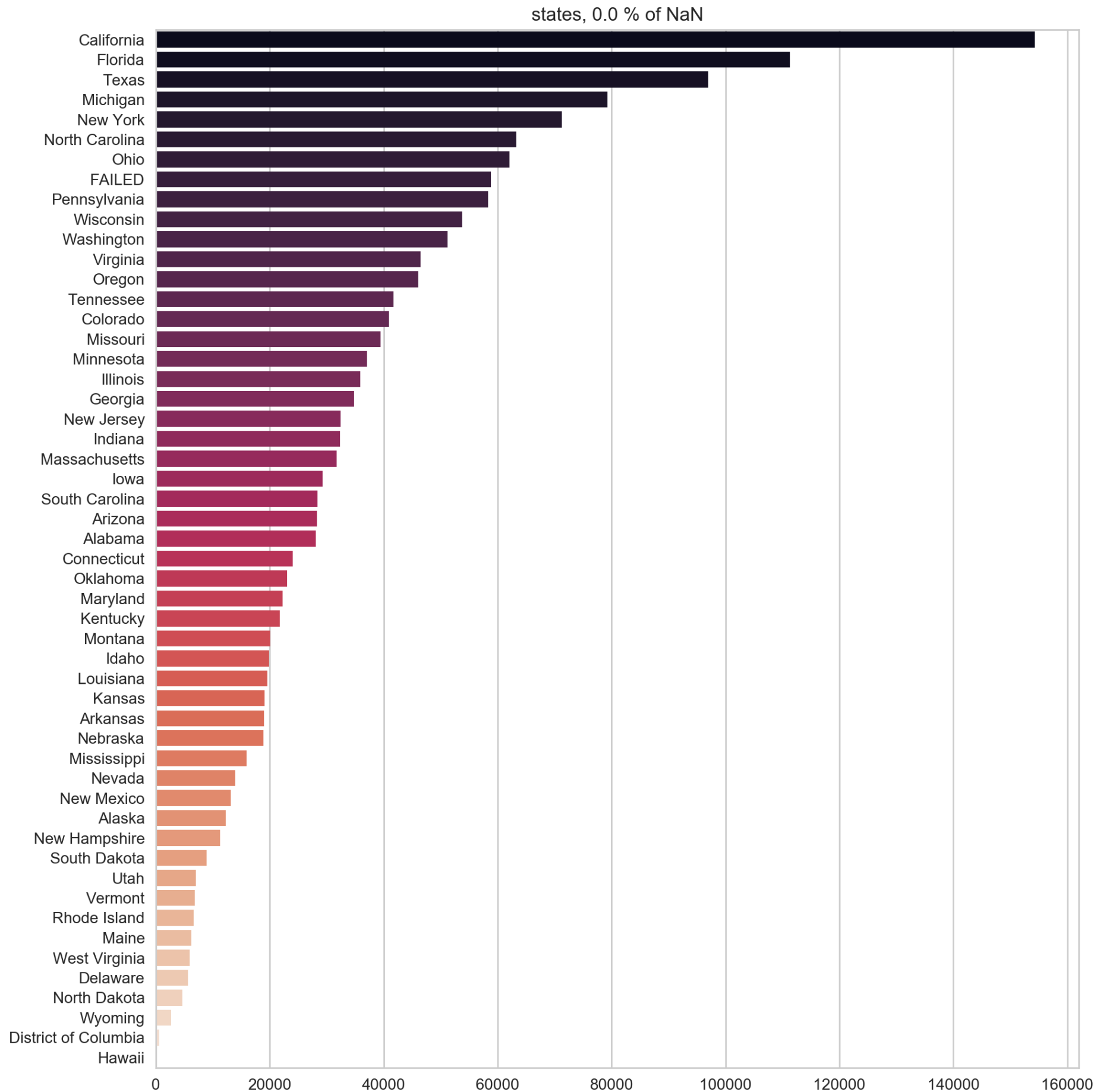
↑ два крайних значения пробега 0 и 10^7

↑ Что за высокие столбики?
(опять круглые значения)



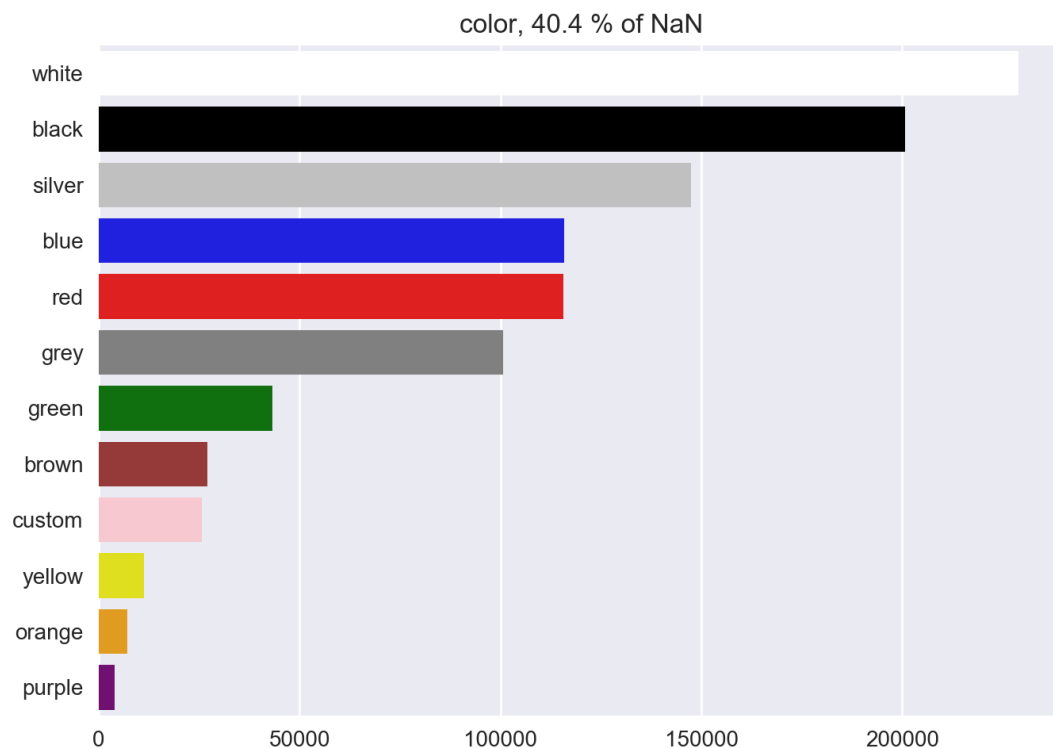
-> два горба (100 и 100000 миль)

Посмотрим на базовые графики



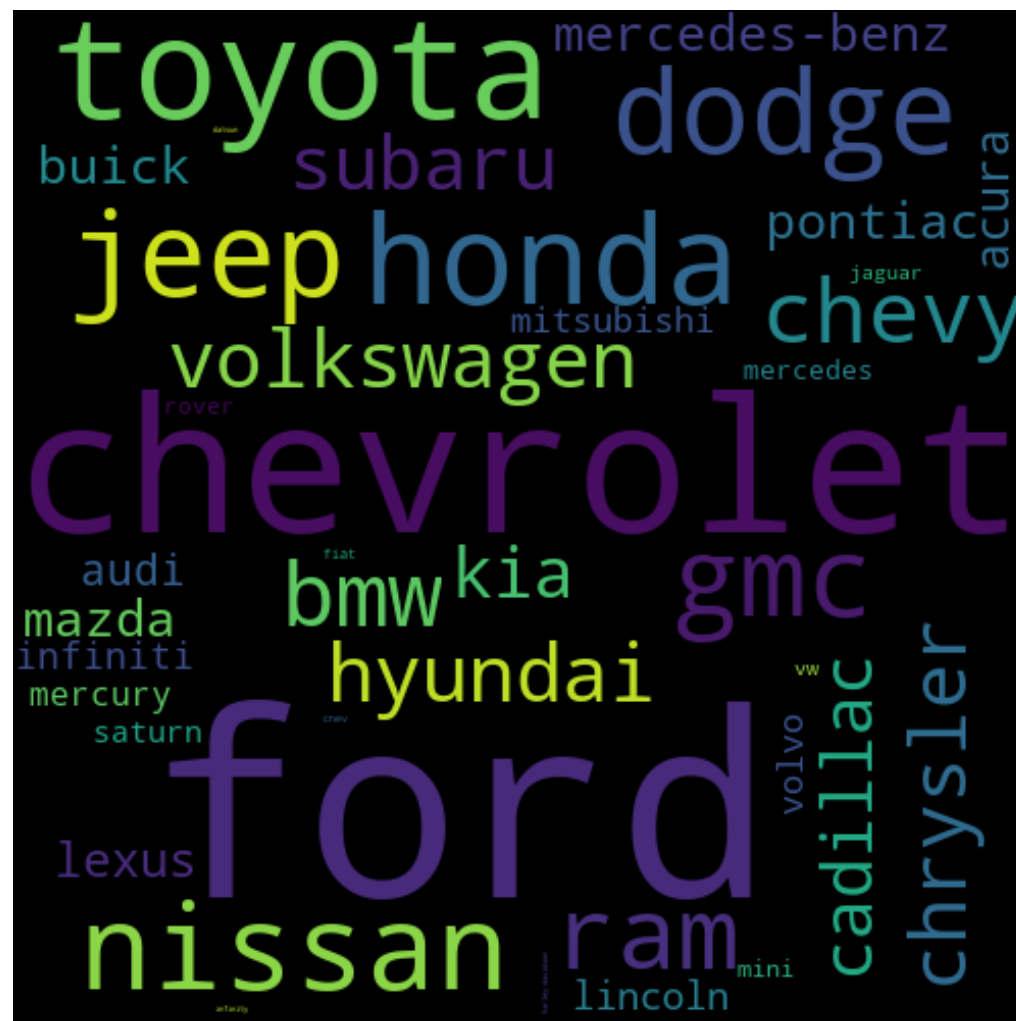
52 “штата”...
Какие лишние?

Посмотрим на базовые графики

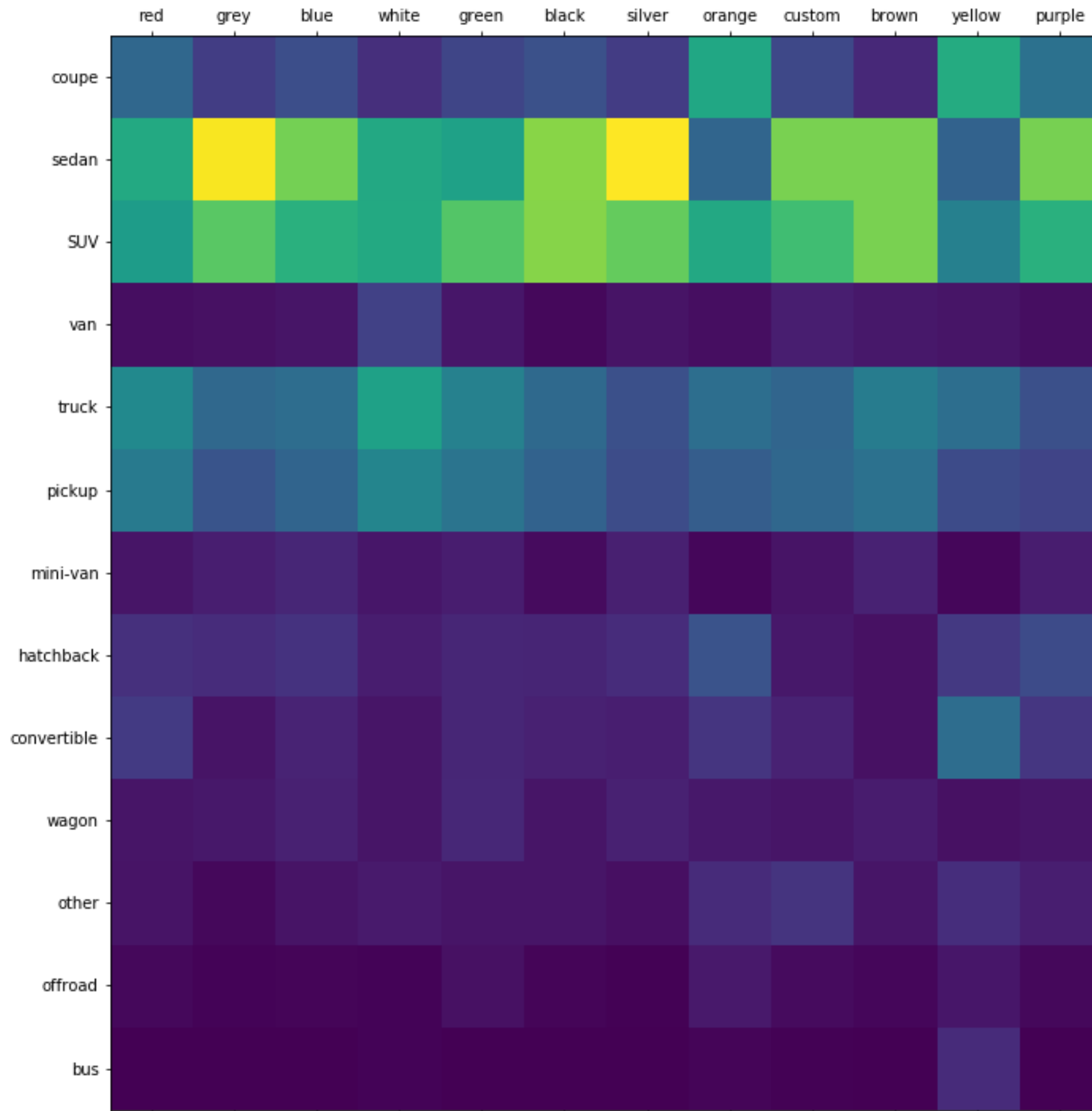


**custom - с рисунком/
несколько цветов**

**В топе американские,
японские и корейские
производители**

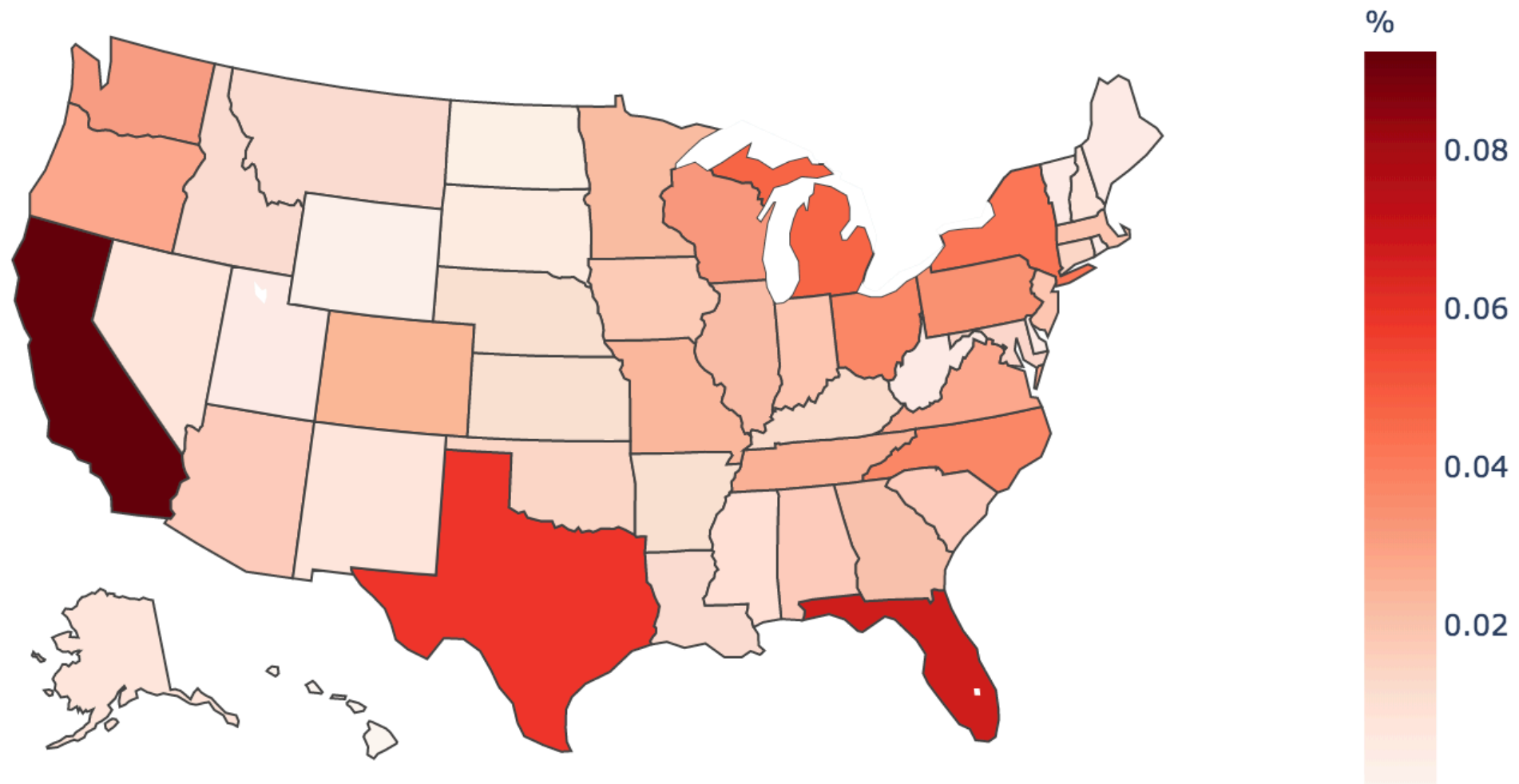


Посмотрим на базовые графики



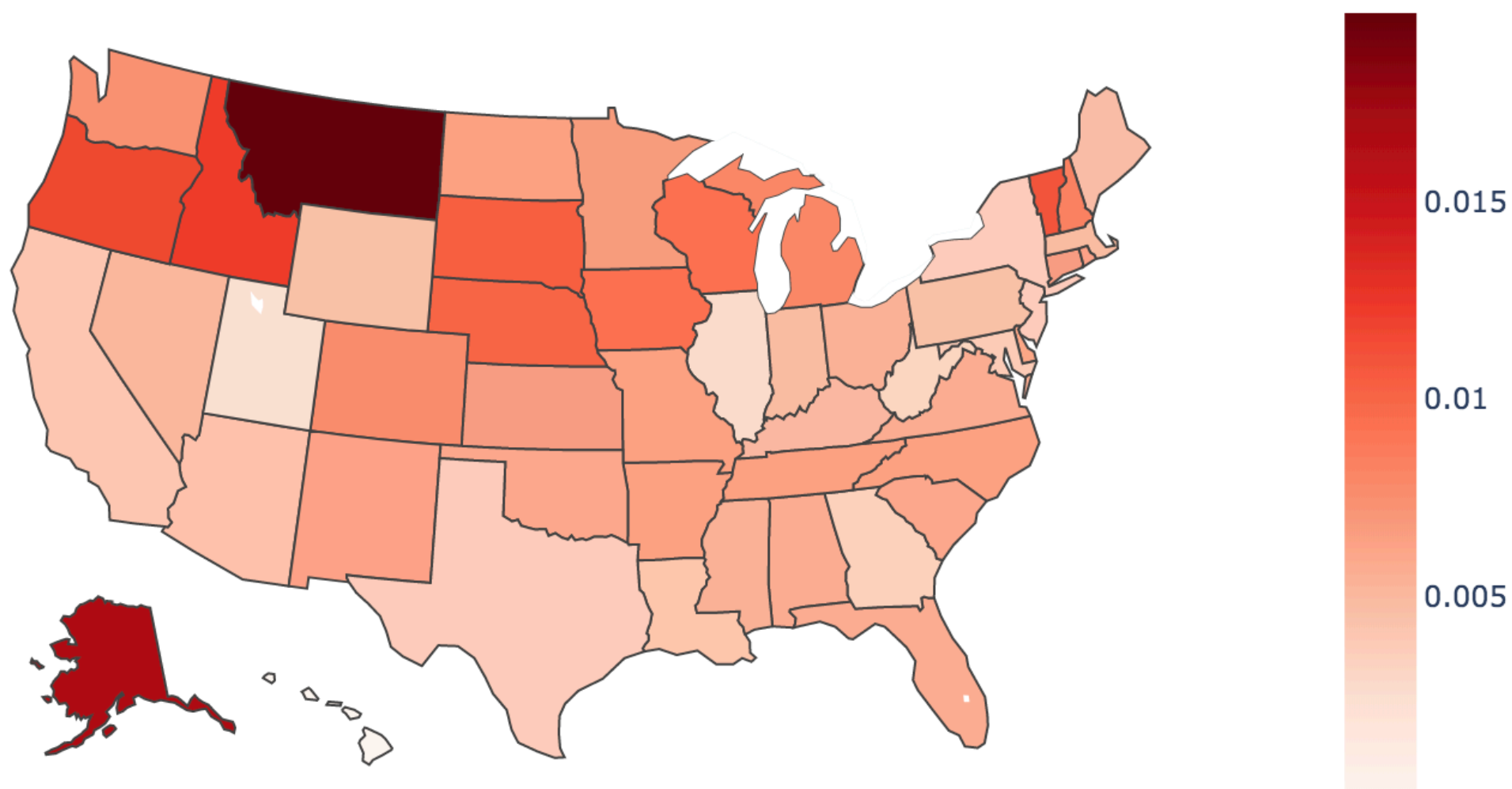
Карта США

По проценту объявлений



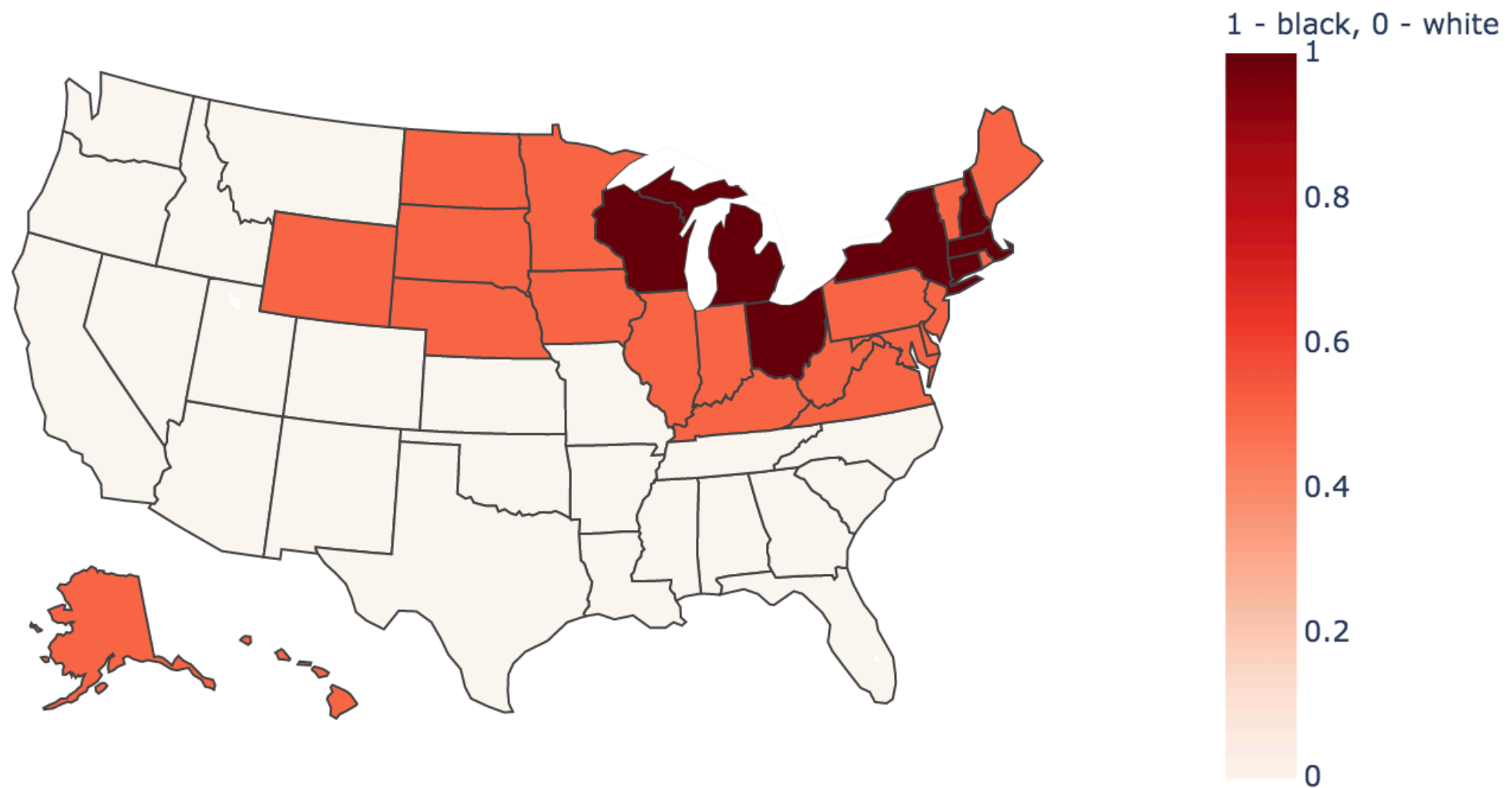
Карта США

Объявление на человека



Карта США

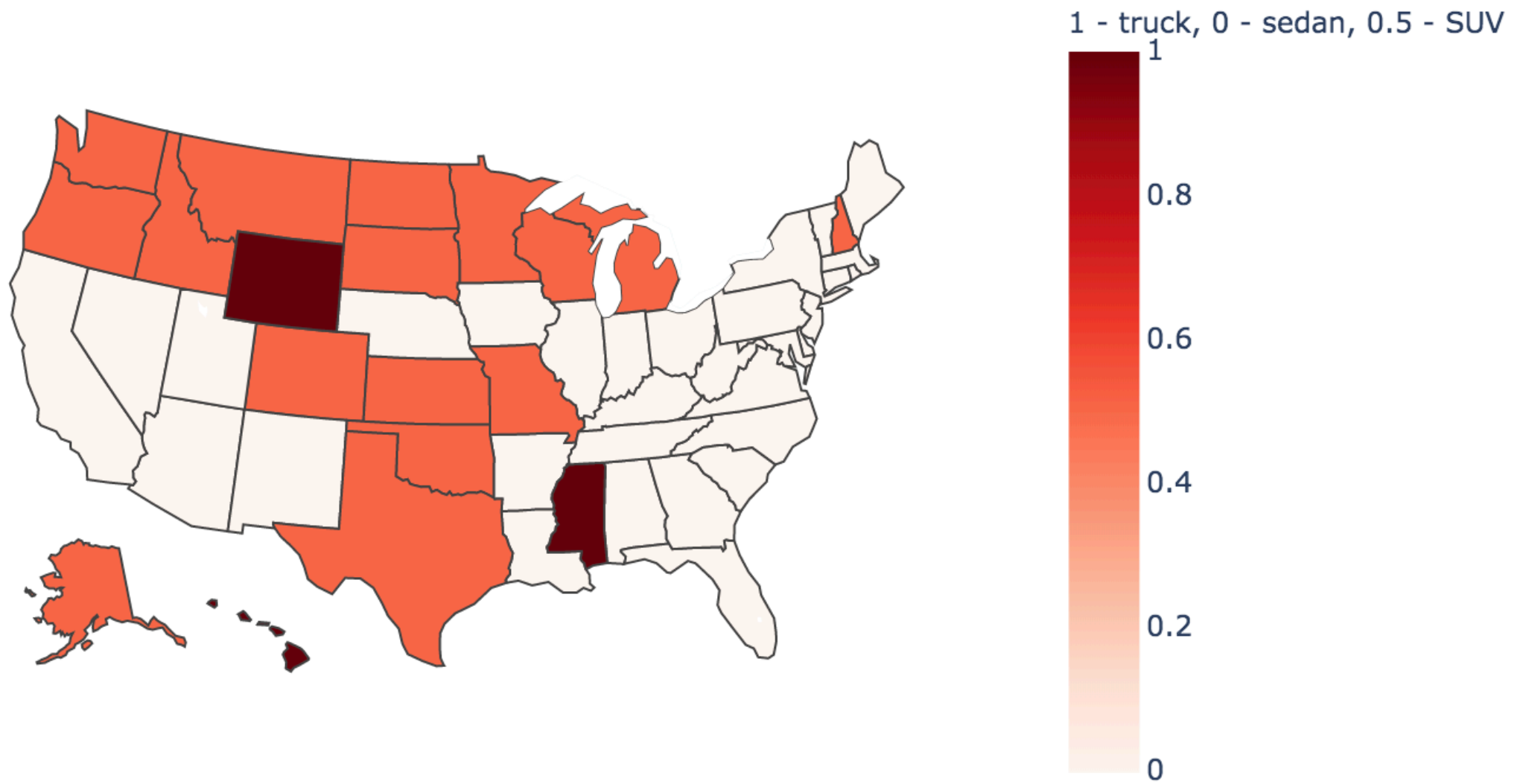
По цветам



**0.5, если разница между первым и
вторым местом не больше 350
объявлений**

Карта США

По типам



Карта США

Электромобили

