

ОСНОВЫ ВИЗУАЛИЗАЦИИ ДАННЫХ С ПОМОЩЬЮ Matplotlib

Андрей Мещеряков

Data Scientist в EPAM

План модуля

- Введение в визуализацию
- Основы matplotlib
- Детали графика
- Визуализация непрерывных данных
- Визуализация категориальных данных

Что такое визуализация

Визуализация данных

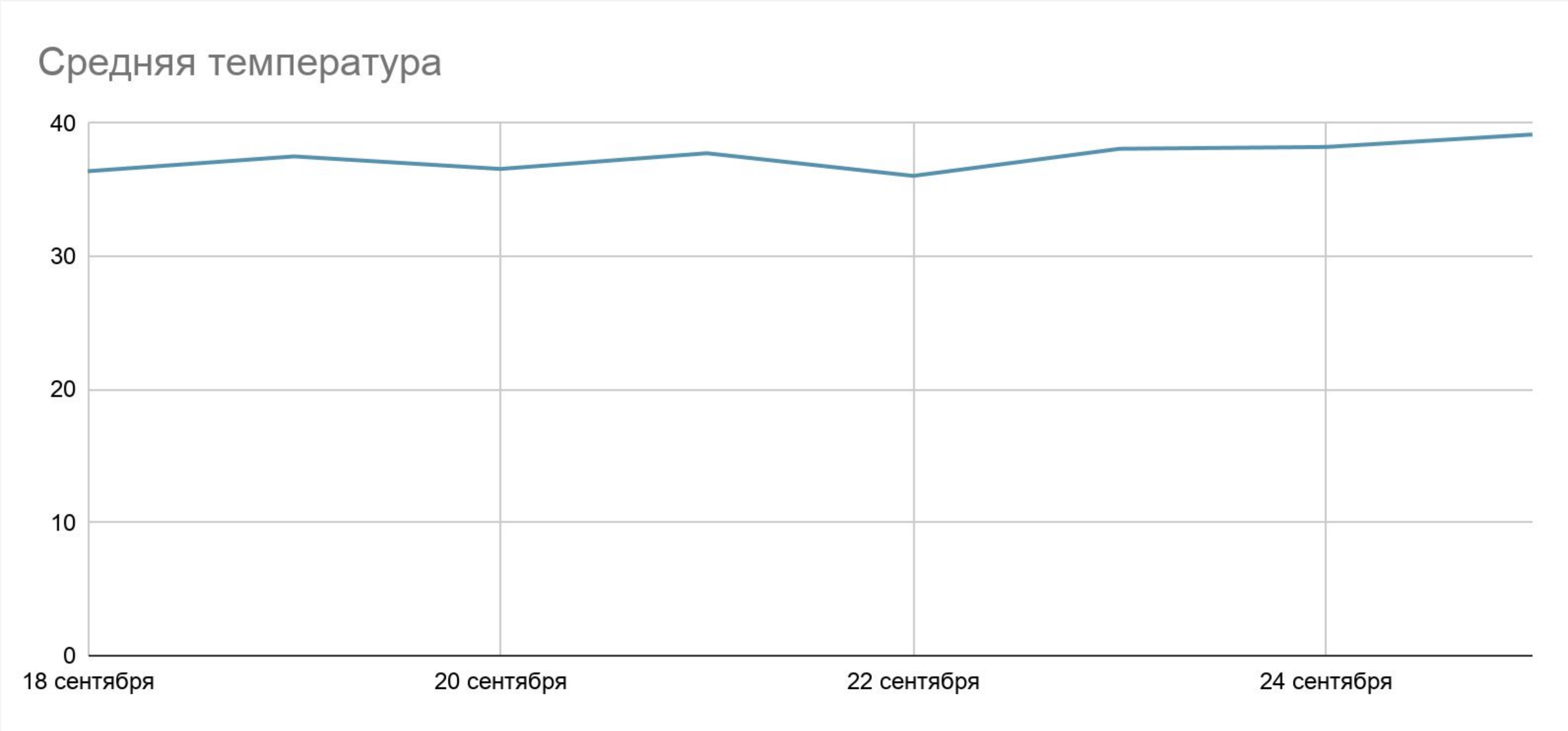
Иллюстрация данных в виде графиков, благодаря которой можно легче увидеть тенденции, закономерности или аномалии в данных

Пример визуализации



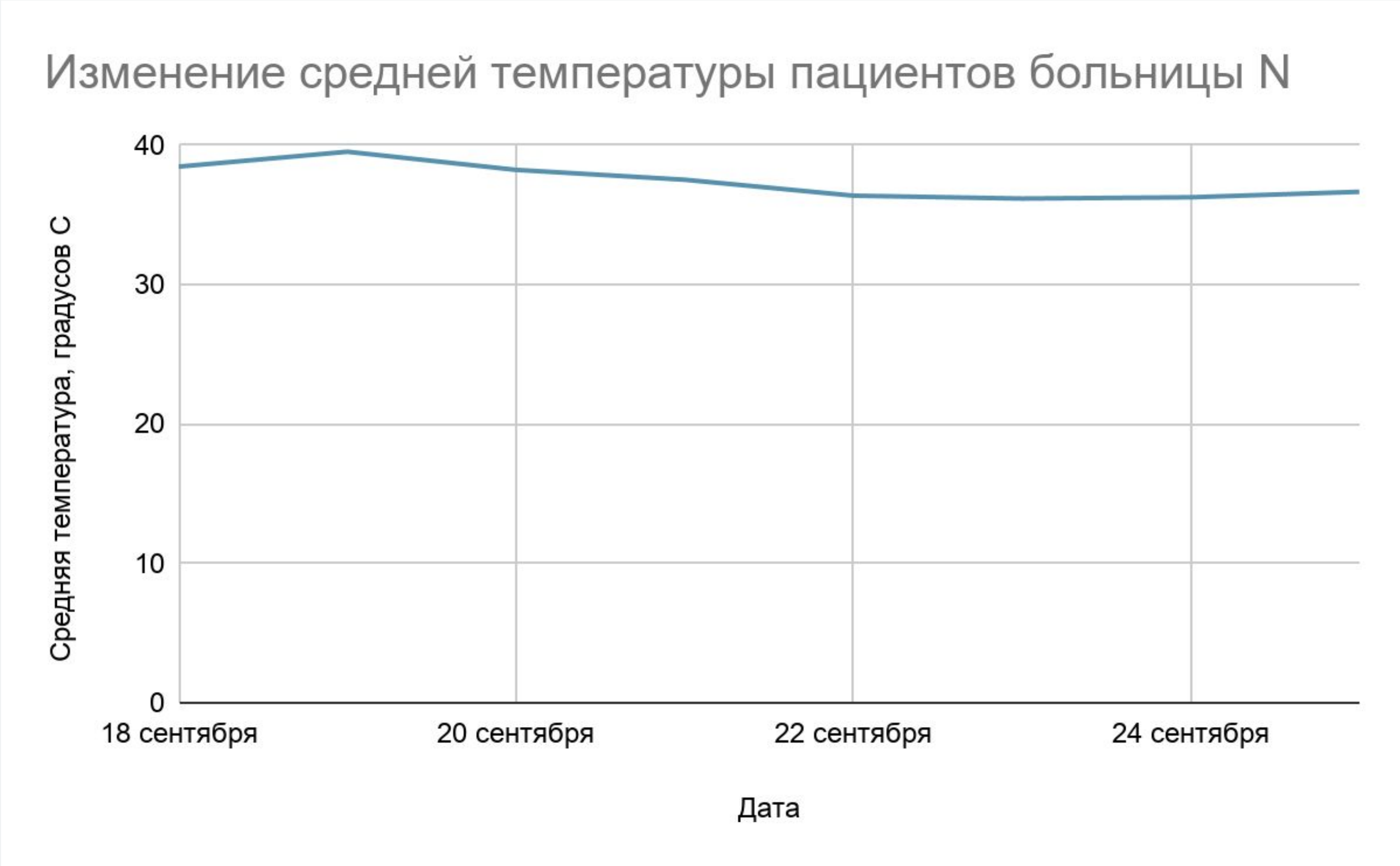
Год	Объем продаж, \$ млн
2012	18
2013	20
2014	19
2015	23
2016	24
2017	36
2018	39
2019	40

Пример визуализации



Пример плохой визуализации – непонятно, какие данные использованы и какое ключевое сообщение несет график

Пример визуализации



Пример плохой визуализации –
бессмысленное ключевое сообщение

Выводы

- Из визуализации должно быть понятно, какие данные использовались для ее построения
- Визуализация должна нести явное и понятное ключевое сообщение
- Ключевое сообщение визуализации должно

Спасибо за внимание!

Знакомство с Matplotlib

Андрей Мещеряков

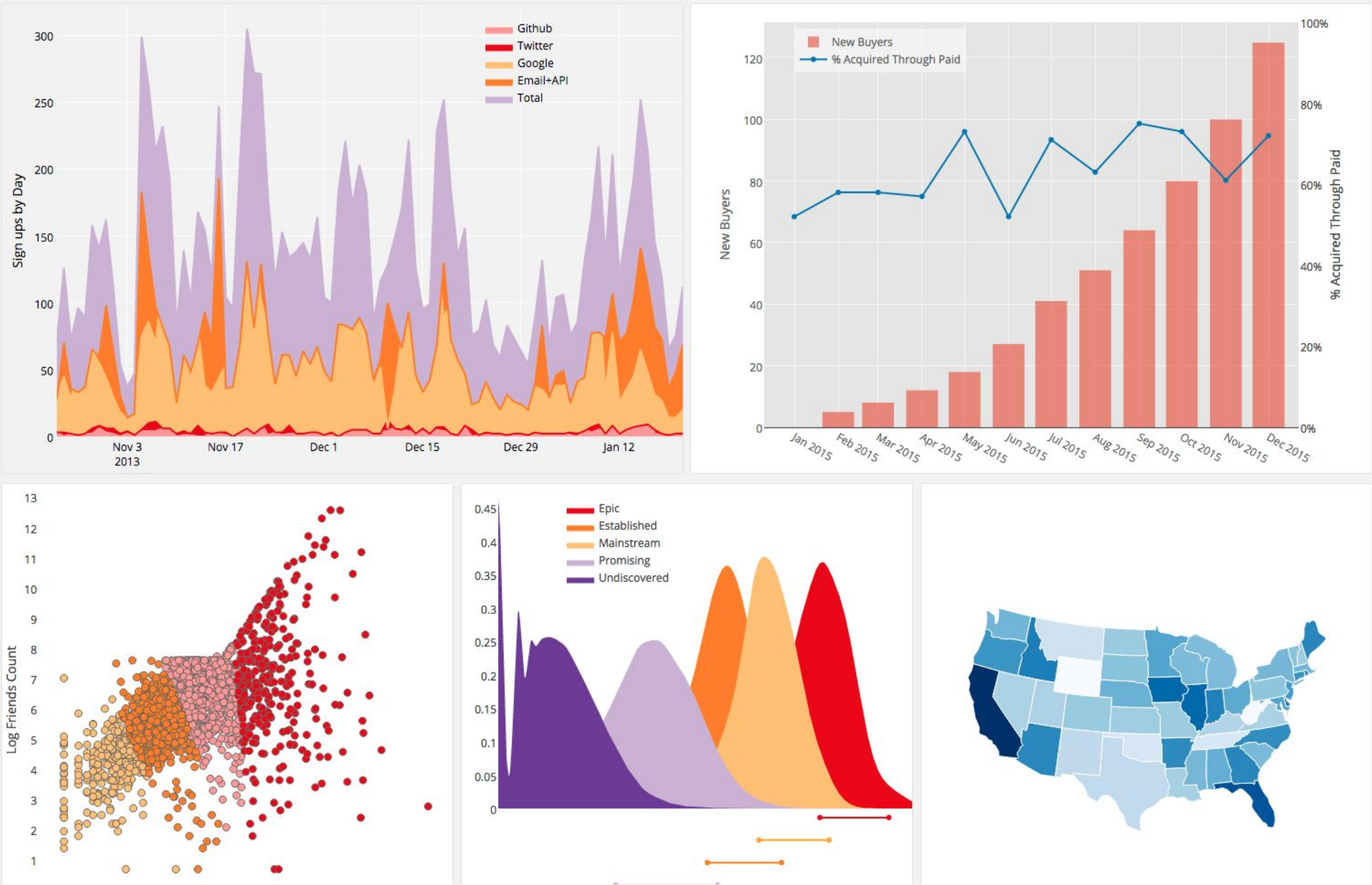
Data Scientist в EPAM

План урока

- Что такое matplotlib
- Как визуализировать данные pandas с помощью matplotlib

Что такое matplotlib

matplotlib – это одна из самых известных и часто используемых Python-библиотек для создания визуализаций



Особенности matplotlib

- matplotlib может визуализировать разные типы данных: списки python, структуры данных numpy, Series и DataFrame pandas и т.д.
- matplotlib - очень гибкая библиотека, которая предлагает большой выбор типов графиков, их настроек, стилей и т.д.

Заглушка для скринкаста

Спасибо за внимание!

Детали графика

Андрей Мещеряков

Data Scientist в EPAM

План урока

- Разбор практического задания
- Добавление деталей графика

Заглушка для скринкаста

Спасибо за внимание!

Визуализация численных данных

Андрей Мещеряков

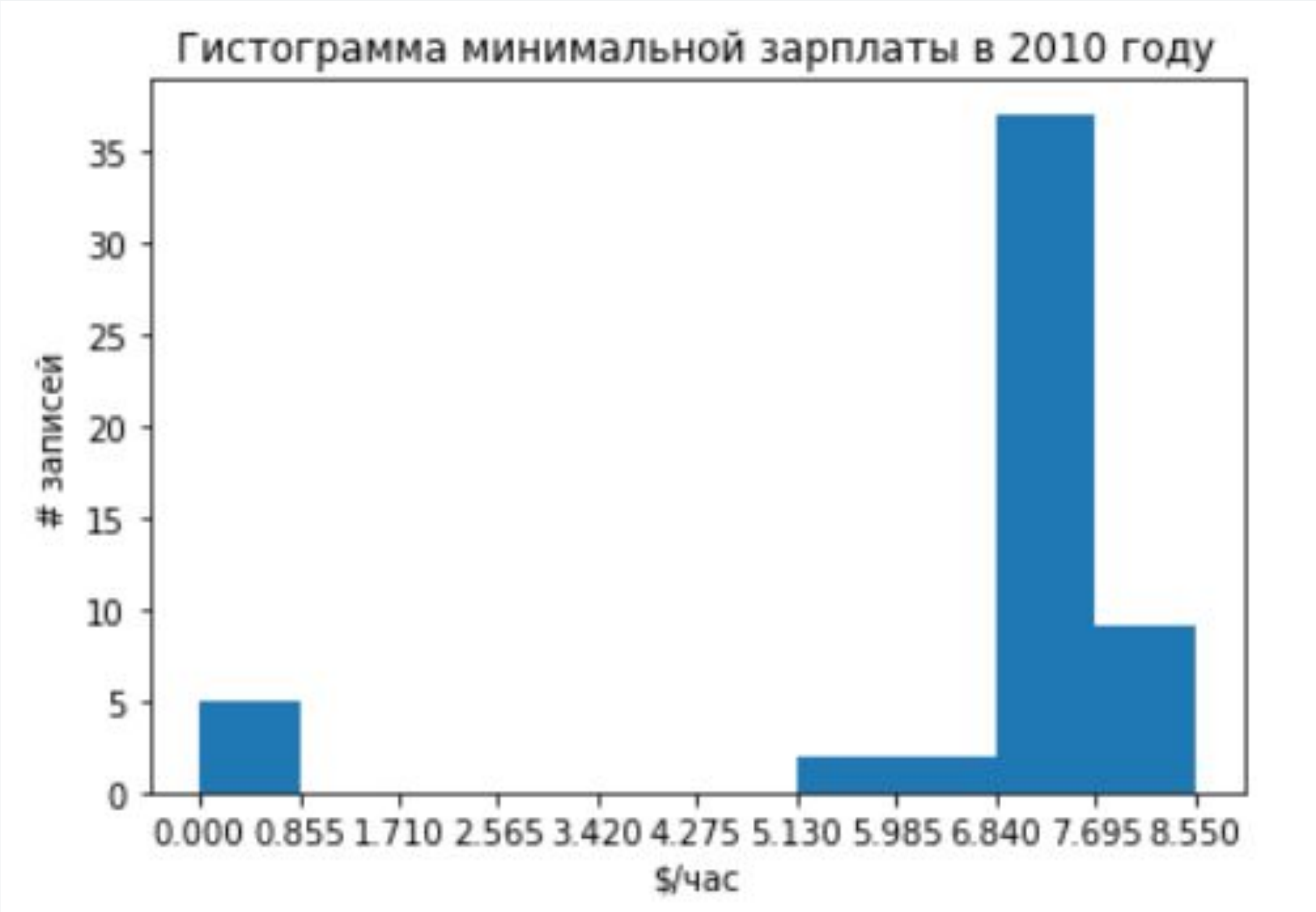
Data Scientist в EPAM

План урока

- Разбор практического задания
- Виды данных
- Типы графиков для визуализации численных данных

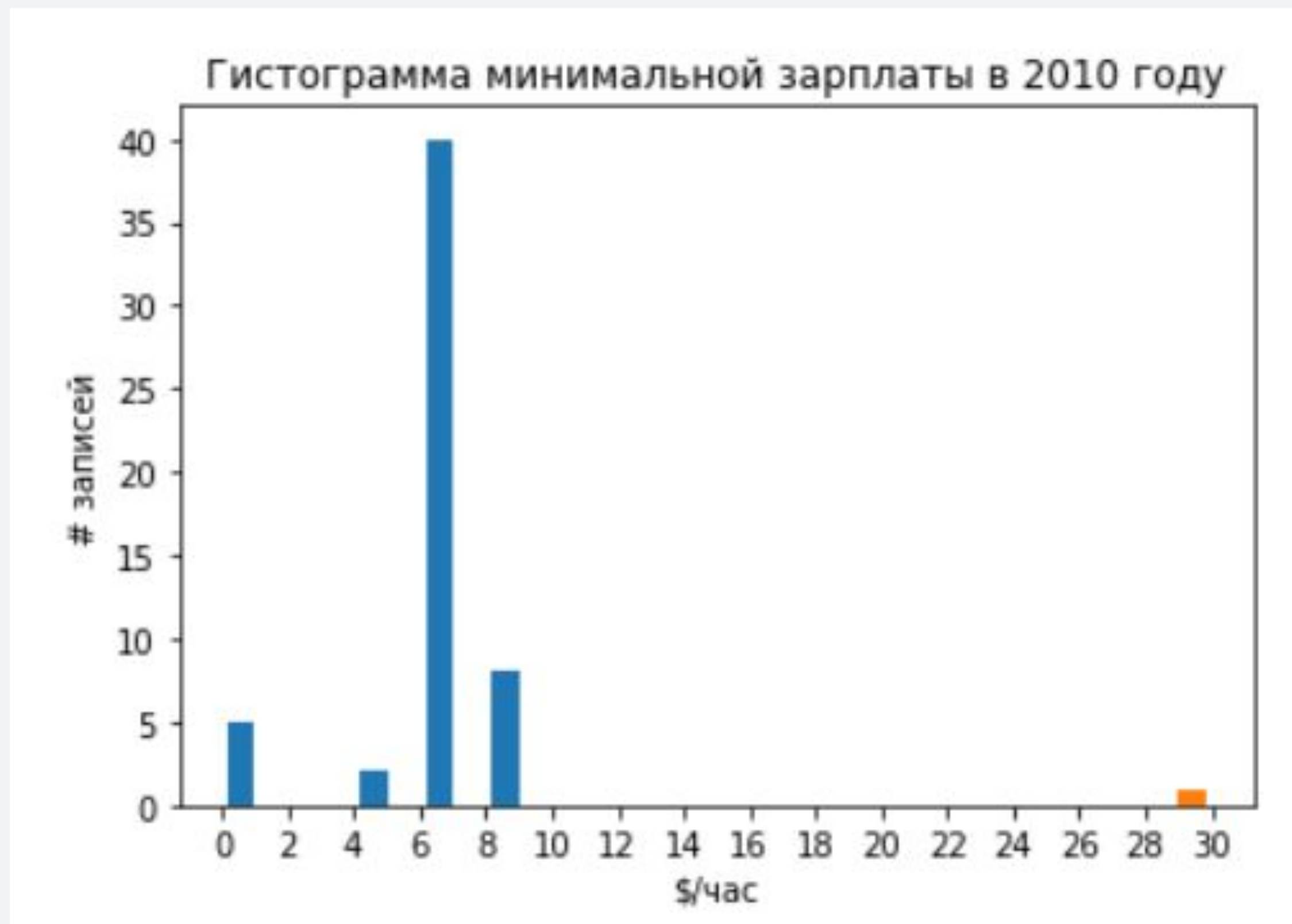
Заглушка для скринкаста

Гистограмма



Гистограмма показывает **распределение** данных

Гистограмма



Гистограмма позволяет быстро обнаружить некоторые **аномалии** данных

Заглушка для скринкаста

Точечный график



Задача точечного графика – показать **зависимость** между признаками на большом количестве точек

Заглушка для скринкаста

Спасибо за внимание!

Визуализация категориальных данных

Андрей Мещеряков

Data Scientist в EPAM

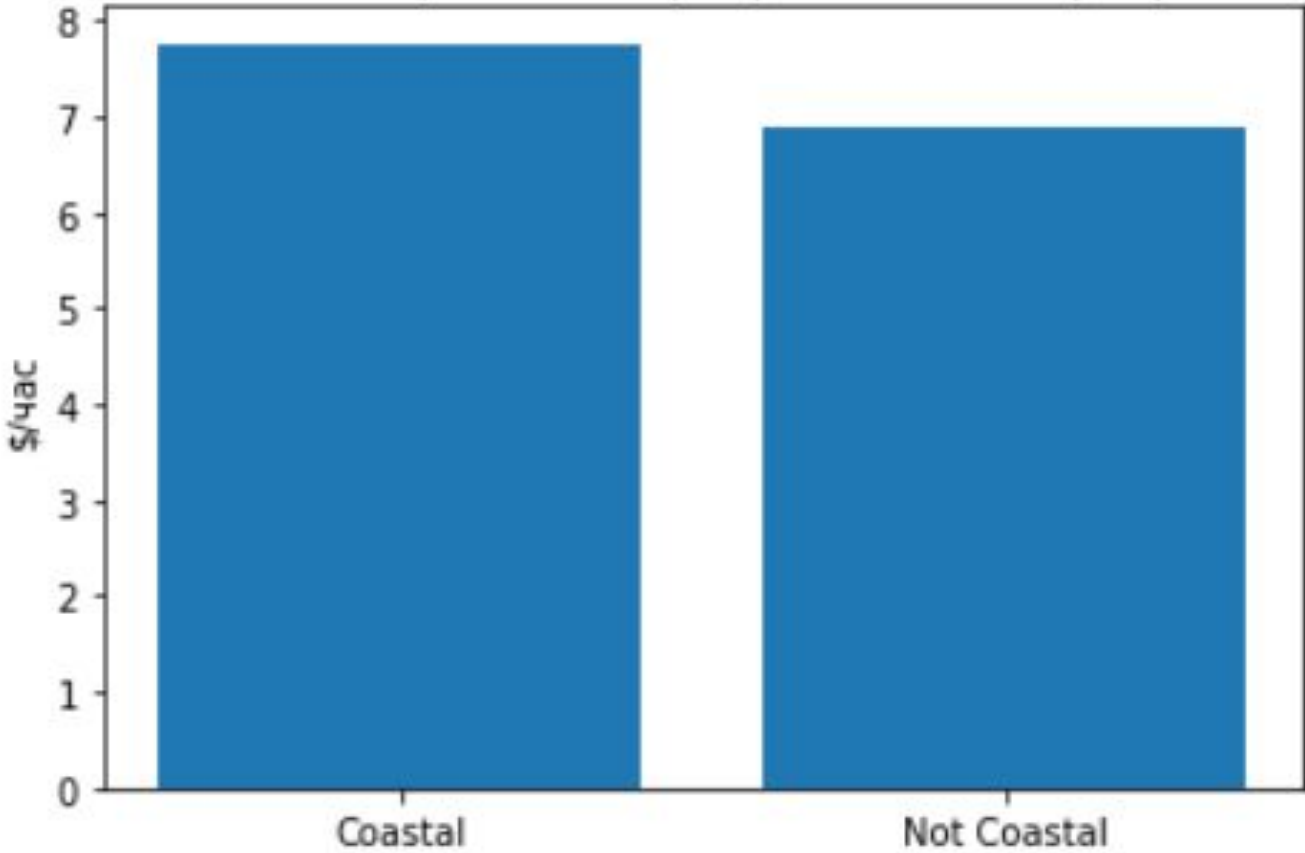
План урока

- Разбор практического задания
- Типы графиков для визуализации категориальных данных

Заглушка для скринкаста

Столбчатая диаграмма

Среднее значение минимальной зарплаты в прибрежных и неприбрежных штатах в 2015 году

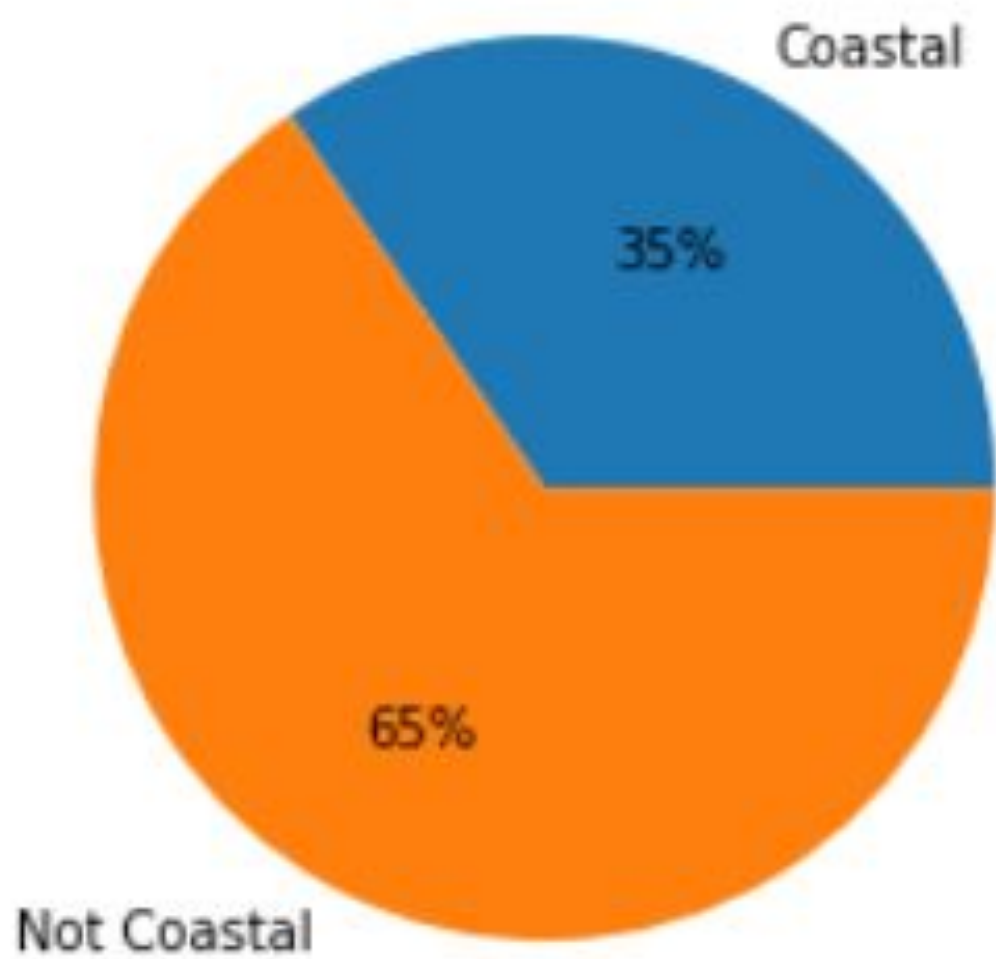


Столбчатая диаграмма используется для **сравнения** разных признаков или категорий

Заглушка для скринкаста

Круговая диаграмма

Соотношение количества прибрежных и неприбрежных штатов



Круговая диаграмма используется для **сравнения частей целого**

Спасибо за внимание!

Частые ошибки при создании визуализации

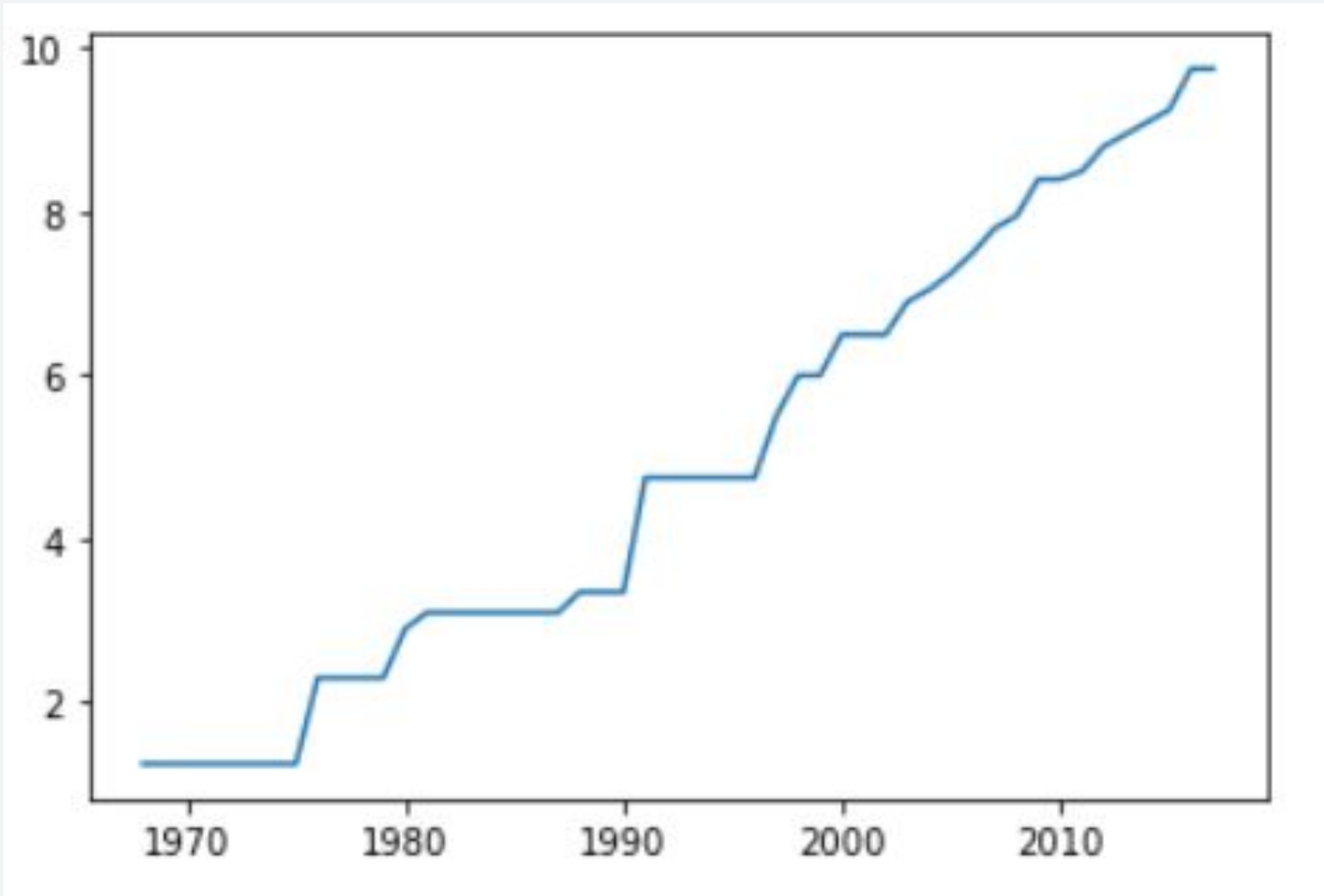
Андрей Мещеряков

Data Scientist в EPAM

План урока

- Разбор практического задания
- Разбор частых ошибок при создании визуализаций

Ошибки при создании визуализации



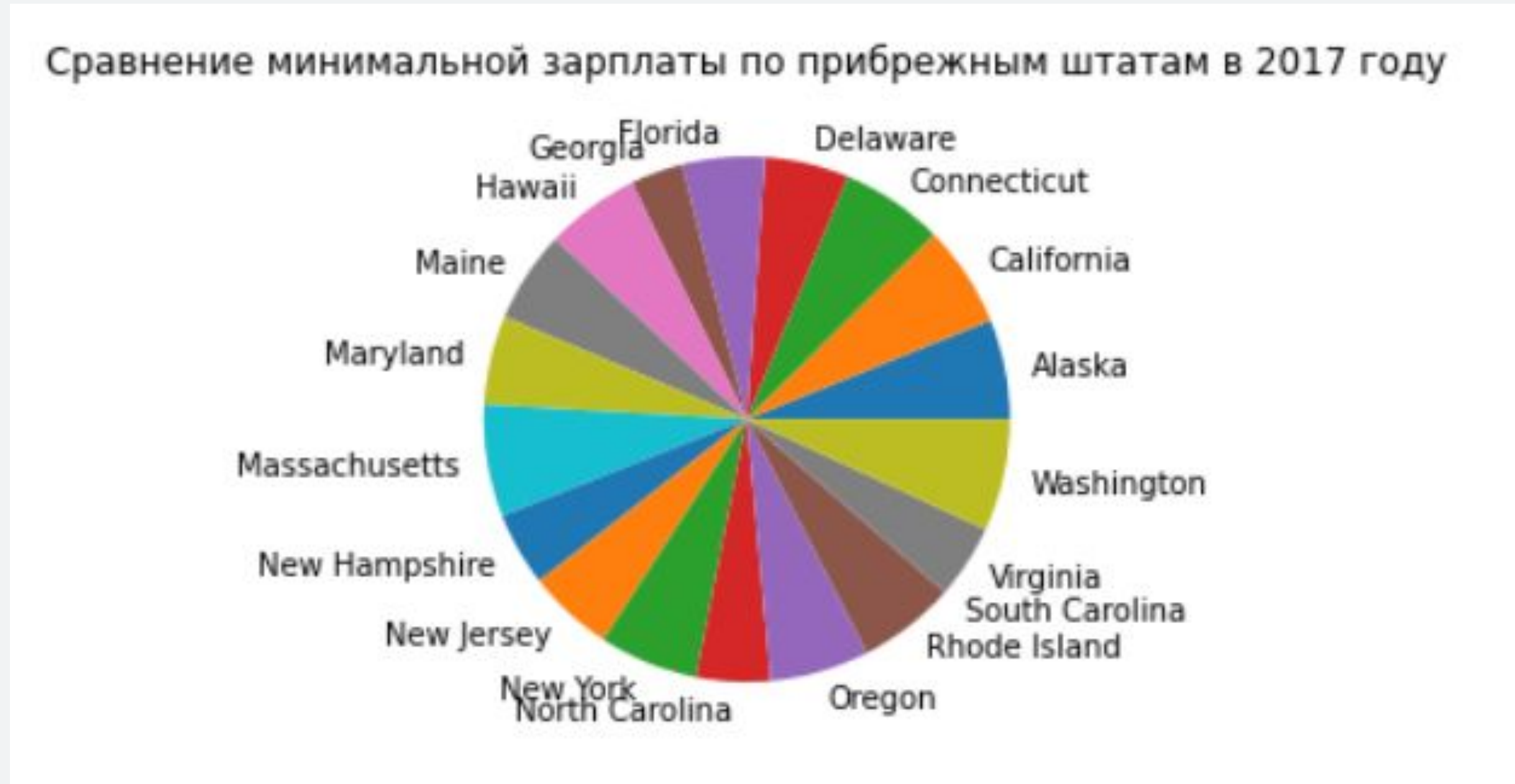
Отсутствующие подписи

Ошибки при создании визуализации



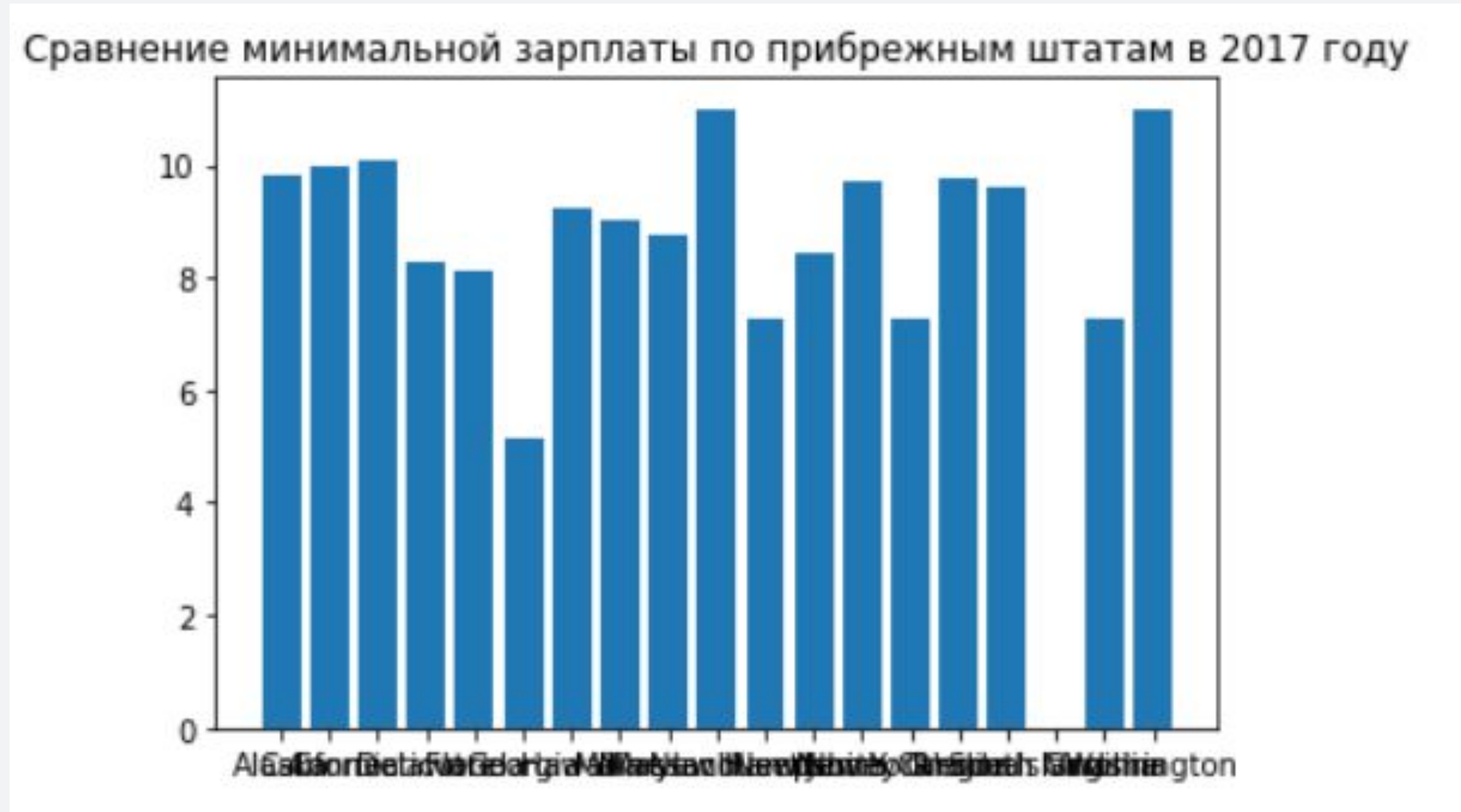
Отсутствующие подписи

Ошибки при создании визуализации



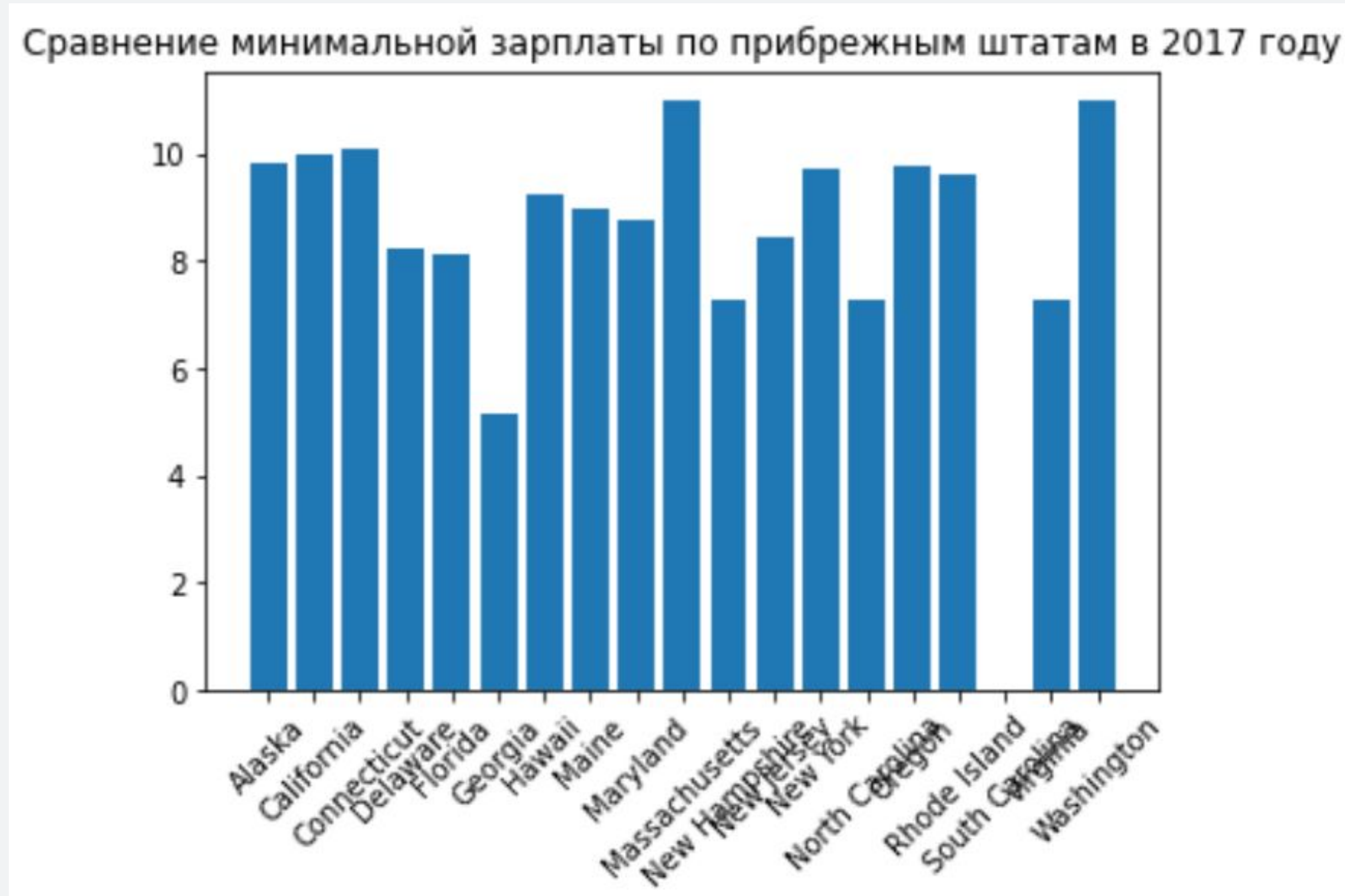
Неверный тип графика

Ошибки при создании визуализации



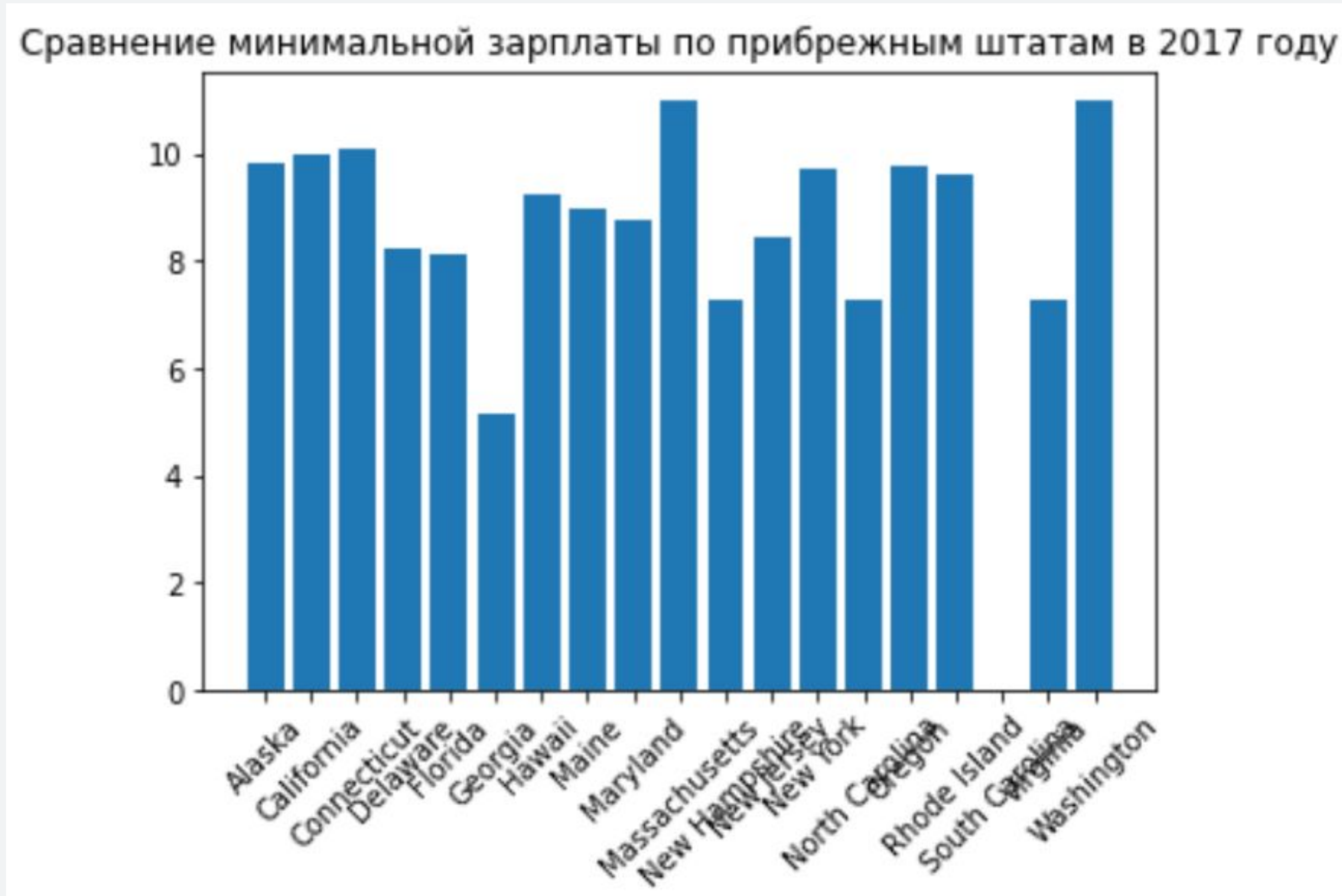
Перегруженный/нечитаемый график

Ошибки при создании визуализации



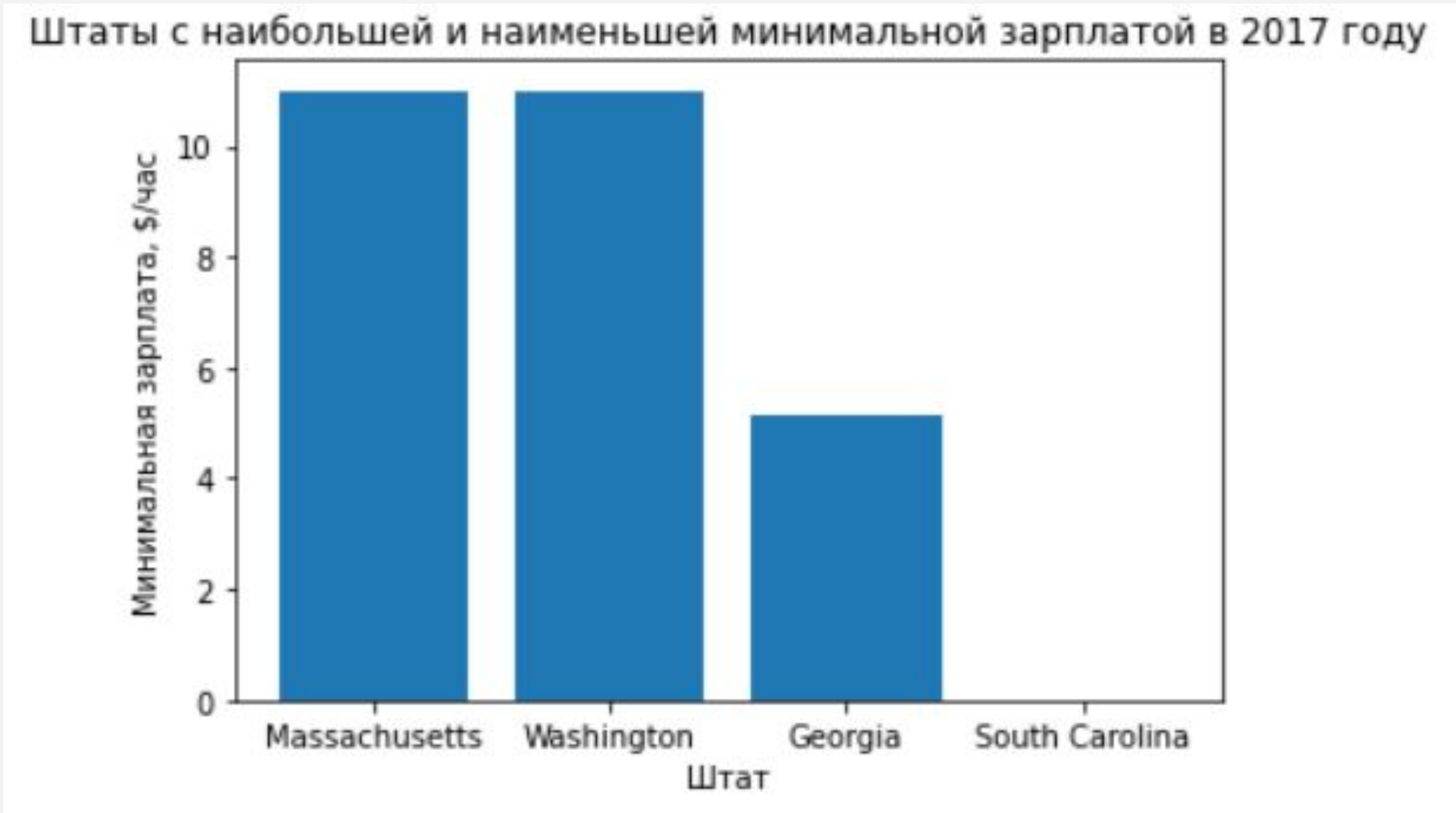
Перегруженный/нечитаемый график

Ошибки при создании визуализации



Невнятное ключевое сообщение

Ошибки при создании визуализации



Невнятное ключевое сообщение

Выводы

- Из визуализации должно быть понятно, какие данные использовались для ее построения
- Визуализация должна нести явное и понятное ключевое сообщение
- Вид графика должен соответствовать типу данных и ключевому сообщению

Спасибо за внимание!