

О датасете - <https://www.kaggle.com/robikscube/hourly-energy-consumption/data>

В датасете представлены данные о почасовом потреблении электроэнергии (МВт) в США

Данные разделены по компаниям, предоставляющим услуги электроснабжения в разных регионах страны

Данные от каждой компании охватывают свой период наблюдений

Однако все они охватывают период с 2013 по 2018 гг.(кроме PJM_load - 1998 - 2002 гг. и NI - 2004 - 2011 гг.)

Структура данных в датасете:

RangelIndex: 178262 entries, 0 to 178261

Data columns (total 13 columns):

Column Non-Null Count Dtype

0 Datetime 178262 non-null object

1 AEP 121273 non-null float64

2 COMED 66497 non-null float64

3 DAYTON 121275 non-null float64

4 DEOK 57739 non-null float64

5 DOM 116189 non-null float64

6 DUQ 119068 non-null float64

7 EKPC 45334 non-null float64

8 FE 62874 non-null float64

9 NI 58450 non-null float64

10 PJME 145366 non-null float64

11 PJMW 143206 non-null float64

12 PJM_Load 32896 non-null float64

Перед началом работы с данными было выдвинуто 4 гипотезы:

- В выходные дни (суббота, воскресенье - 5,6 дни недели) среднее энергопотребление падает по сравнению с рабочей неделей (+)
- Среднесуточное энергопотребление в летний период ниже, чем в другие времена года (-)
- Энергопотребление в среднем растет из года в год (-)
- Пик энергопотребления приходится на время после окончания рабочего дня (17:00 - 23:00) (+)

Затем была начата подготовка данных. В первую очередь было выяснено, что представленные метрики описывают один и тот же часовой пояс.

Сначала проверим гипотезу об изменении тенденции энергопотребления из года в год. Для этого найдем сумму наших метрик, описывающих один регион, находящийся в одном часовом поясе. Запишем результат в новый столбец energy_total_east. Найдем среднее энергопотребление в час по каждому дню. Эта характеристика будет основной на нашем первом графике.

Среднее энергопотребление восточного региона США в час по дням. Июль 2013 - Август 2018гг.

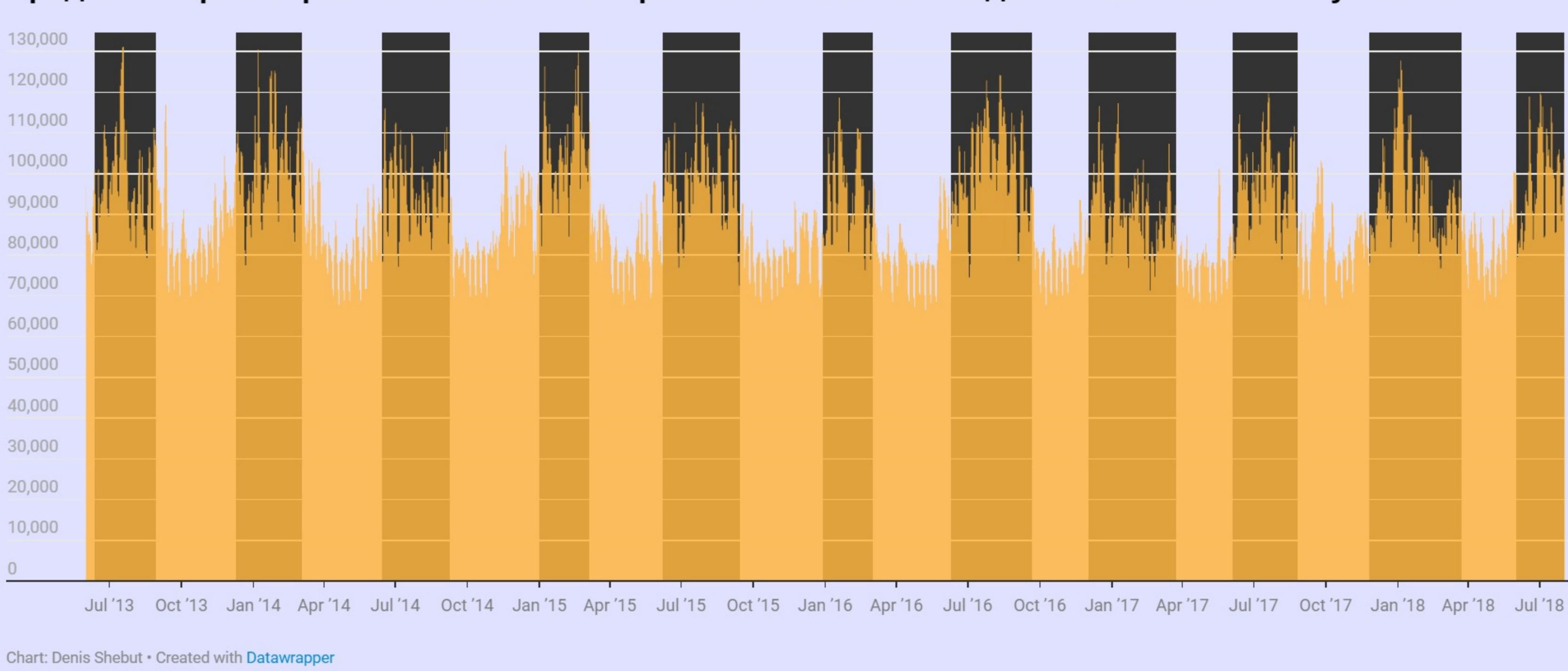


Chart: Denis Shebut • Created with Datawrapper

Максимумы энергопотребления приходятся на:

(-1-) 2013/06/12---2013/08/29 (~79 дней)

(-2-) 2013/12/10---2014/03/04 (~85 дней)

(-3-) 2014/06/14---2014/09/09 (~88 дней)

(-4-) 2015/01/01---2015/03/06 (~65 дней)

(-5-) 2015/06/08---2015/09/14 (~99 дней)

(-6-) 2015/12/29---2016/03/02 (~65 дней)

(-7-) 2016/06/10---2016/09/21 (~104 дней)

(-8-) 2016/12/02---2017/03/24 (~113 дней)

(-9-) 2017/06/04---2017/08/26 (~84 дней)

(-10-) 2017/11/26---2018/03/23 (~118 дней)

(-11-) 2018/06/01---2018/08/01 (~62 дней - конец периода наблюдений)

Количественного роста энергопотребления не наблюдается. Можно заметить некоторое увеличение длительности максимумов. Однако, как мне кажется, так как разметка промежутков производилась вручную, сказать что-то определенное об изменении их длительности нельзя. Изменение находится в пределах погрешности.

Чтобы проверить гипотезу №2, отобразим второй график, на котором среднее энергопотребление в час по дням в разные года можно сопоставить более наглядно. Также выделим промежутками времена года.

Среднее энергопотребление восточного региона США в час по дням. Июль 2013 - Август 2018гг.

Сопоставление данных нескольких лет. Цветами выделены времена года.

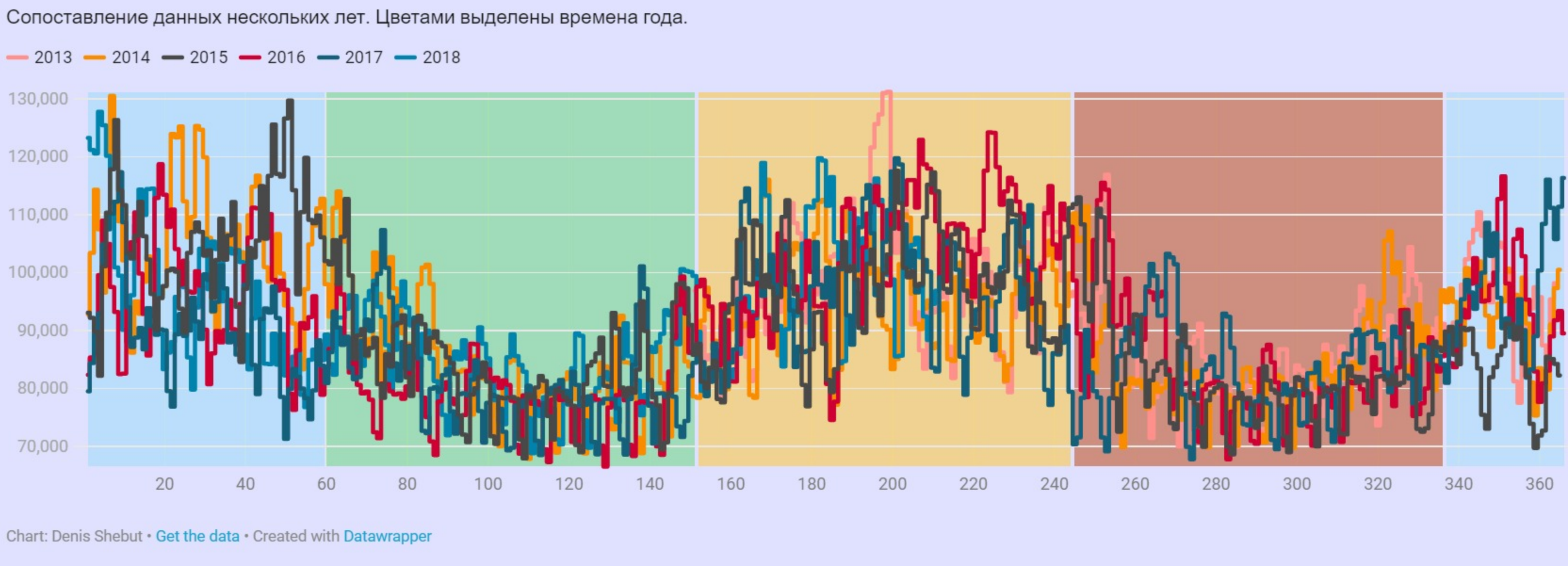
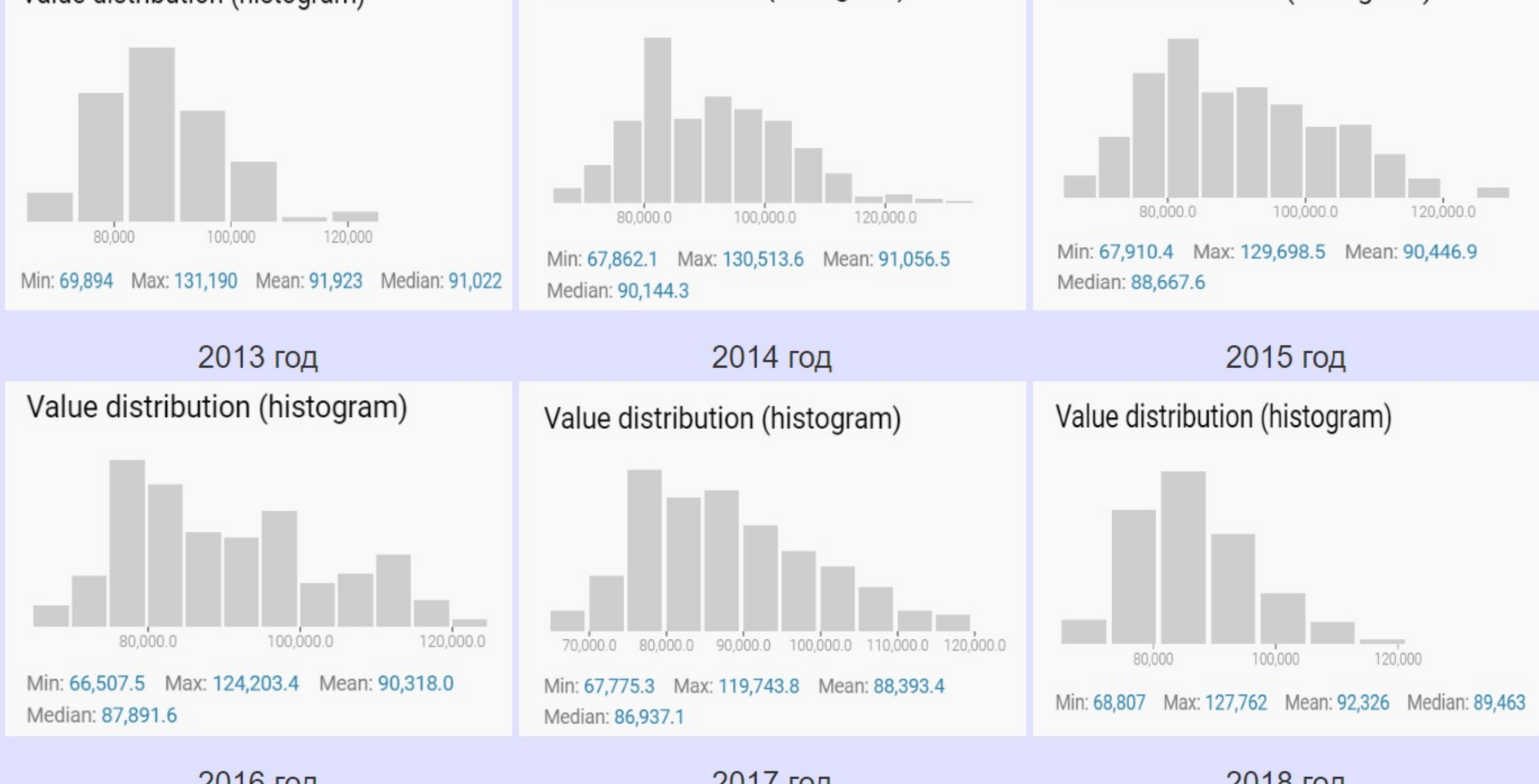


Chart: Denis Shebut • Get the data • Created with Datawrapper

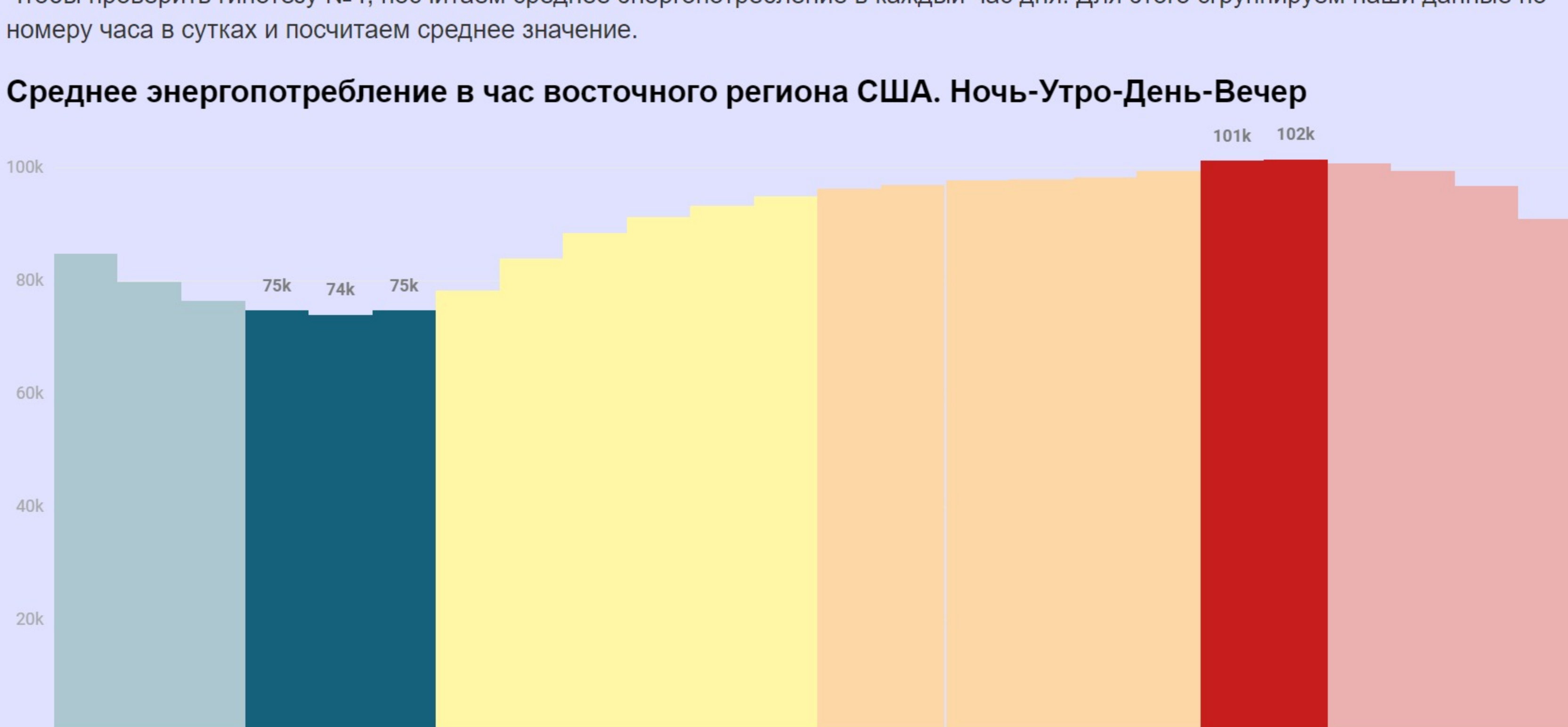
Распределение значений в исходных данных по годам:



Занимательный факт. Если взглянуть внимательнее на распределение значений, то видно, что 2016 год выделяется отклонением в сторону больших значений (для 2013 и 2018 годов предоставлены данные только по первому и второму полугодью соответственно). Предположительно, такое отклонение вызвано тем, что в какие-то определенные дни энергопотребление резко возросло и таких дней было больше чем обычно. Можно попробовать объяснить данное отклонение президентскими выборами в США в 2016 году.

Чтобы проверить гипотезу №4, посчитаем среднее энергопотребление в каждый час дня. Для этого сгруппируем наши данные по номеру часа в сутках и посчитаем среднее значение.

Среднее энергопотребление в час восточного региона США. Ночь-Утро-День-Вечер



Created with Datawrapper

Отчетливо видно, что максимум энергопотребления приходится на 18-19 часов вечера, а минимум на 3-5 часов ночи.

Теперь, чтобы проверить гипотезу №1, проведем те же манипуляции, но сгруппировав исходные данные по дню недели:

Среднее энергопотребление в час восточного региона США. По дням недели

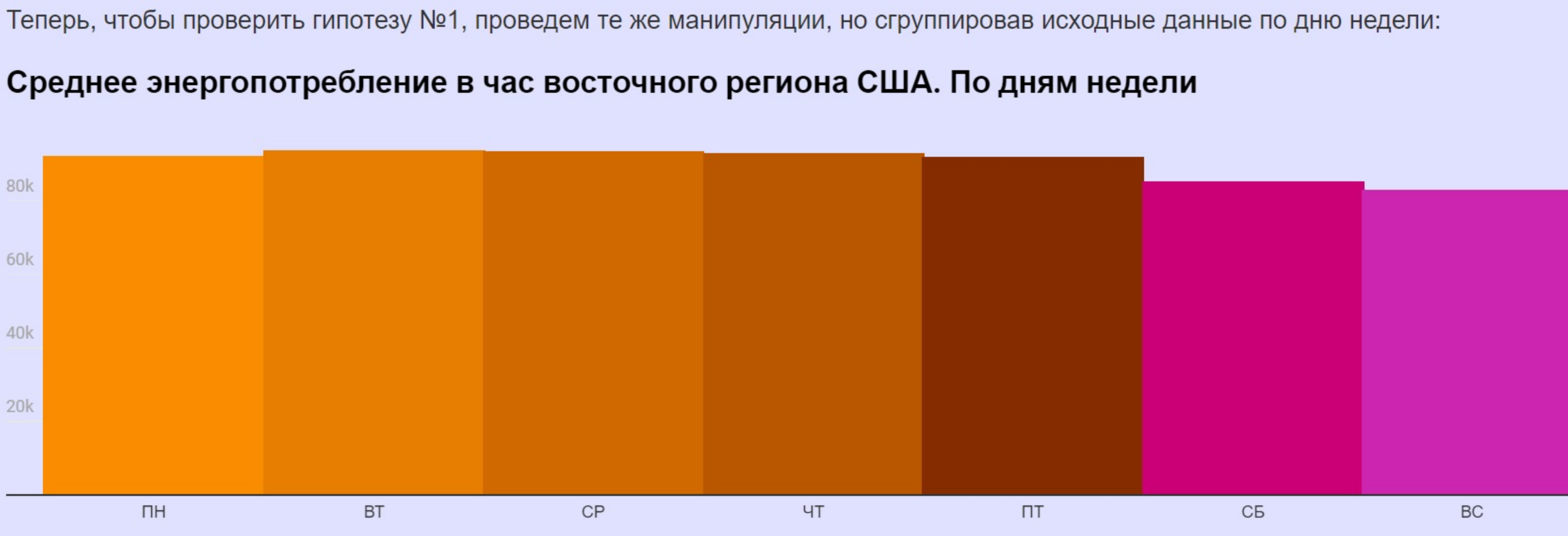


Chart: Denis Shebut • Get the data • Created with Datawrapper

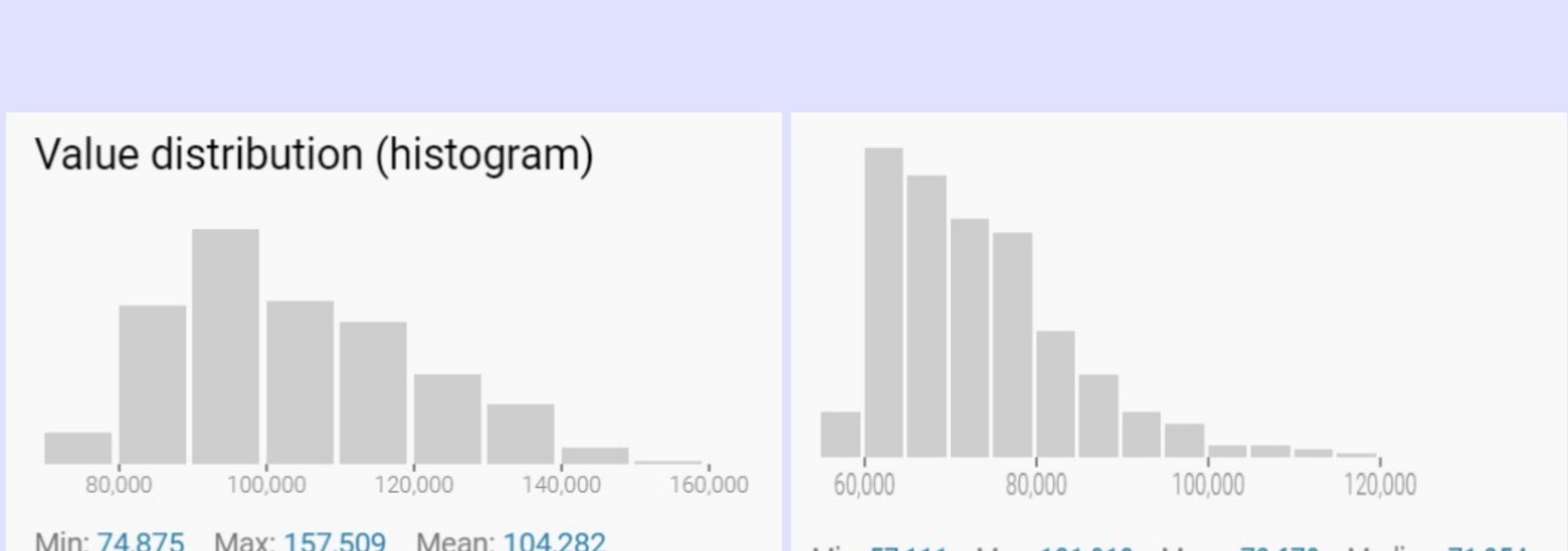
По столбчатой диаграмме видно, что среднее энергопотребление в час в выходные дни меньше чем в будние дни

Чтобы получить более полное представление об исходных данных и выявить величину колебаний энергопотребления в течение дня, построим дополнительно пятый график. Разобьем исходные данные по годам, сгруппируем по номеру дня в году и найдем минимальное и максимальное значения электропотребления в час. Затем склеим данные и отобразим на графике:



Видно, что в летний период разница между дневными максимумами и минимумами наибольшая. Учитывая другие данные, можно предположить, что в летний период в дневное время по рабочим дням энергопотребление значительно растет.

Chart: Denis Shebut • Created with Datawrapper



Распределение значений максимумов

Распределение значений минимумов

Из этого графика мы можем узнать еще одну особенность исходных данных. Видно, что наибольшая разница между максимумами и минимумами наблюдается в (+-) летний период. Учитывая ранее выясненные закономерности, можно предположить, что в теплый период года дневное энергопотребление в будние дни заметно выше чем в остальное время года.