ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ НИЖЕГОРОДСКИЙ
ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Р. Е. АЛЕКСЕЕВА

Кафедра «Прикладная математика»

Лабораторная работа №2

по дисциплине «Базы данных»

Тема: «Визуализация данных на python с использованием pandas и matplotlib»

Студент

_____ Валькова.Н.П.
(Подпись) (Фамилия, И., О.)

18-ПМ ……
(Группа) (Дата сдачи)

Проверил

_____ Моисеев А.Е
(Подпись) (Фамилия, И., О.)

Отчет защищен «___» _____2021_г.
с оценкой_____

Нижний Новгород, 2021

Оглавление

Введение

Matplotlib— это **библиотека** 2D-графиков **Python**, которая выдает показатели качества публикаций в различных печатных форматах и интерактивных средах на разных платформах.
В документации автор признаётся, что *Matplotlib* начинался с подражания графическим командам **MATLAB**, но является независимым от него проектом.
Библиотека *Matplotlib* построена на принципах ООП, но имеет процедурный интерфейс pylab, который предоставляет аналоги команд MATLAB.

В настоящее время пакет работает с несколькими графическими библиотеками, включая WxWindows и PyGTK.
Пакет поддерживает многие виды графиков и диаграмм:
Графики (line plot)

- Диаграммы разброса (scatter plot)
- Столбчатые диаграммы (b»ar chart) и гистограммы (histogram)
- Круговые диаграммы (pie chart)
- Ствол-лист диаграммы (stem plot)
- Контурные графики (contour plot)
- Поля градиентов (quiver)
- Спектральные диаграммы (spectrogram)

Пользователь может указать оси координат, решетку, добавить надписи и пояснения, использовать логарифмическую шкалу или полярные координаты.

Seaborn - это библиотека для визуализации данных и выделения их статистических особенностей. Seaborn написанна поверх библиотеки Matplotlib.
Еще одна важная особенность библиотеки Seaborn - это заложенный в нее механизм предобработки данных, возможный благодаря тесной интеграции с библиотекой Pandas.

*Пояснение к графикам : столбцы отображают усредненное значение зарплаты, верх линии максимальное значение, низ — минимальное.

Постановка задачи:

Проанализировать датасет с сайта ilostst.olo.org используя визуализацию данных.

Для разбора был выбран датасет с средней номинальной ежемесячной заработной платой работников в разбивке по полу и роду занятий.

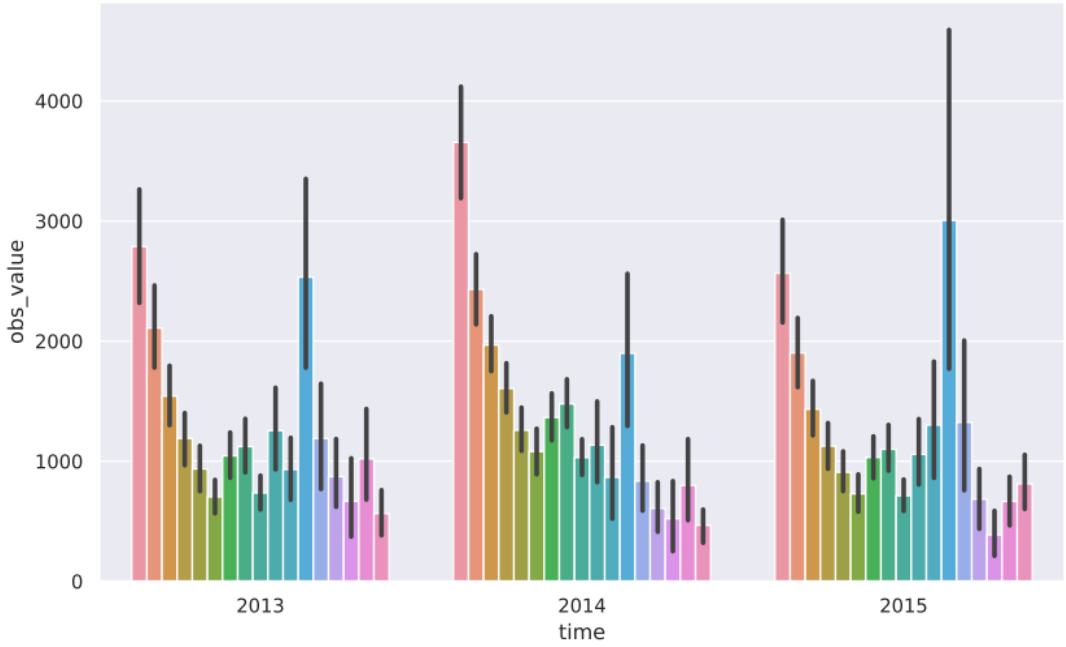
Столбцы отображают усредненное значение зарплаты, верх линии максимальное значение, низ — минимальное.

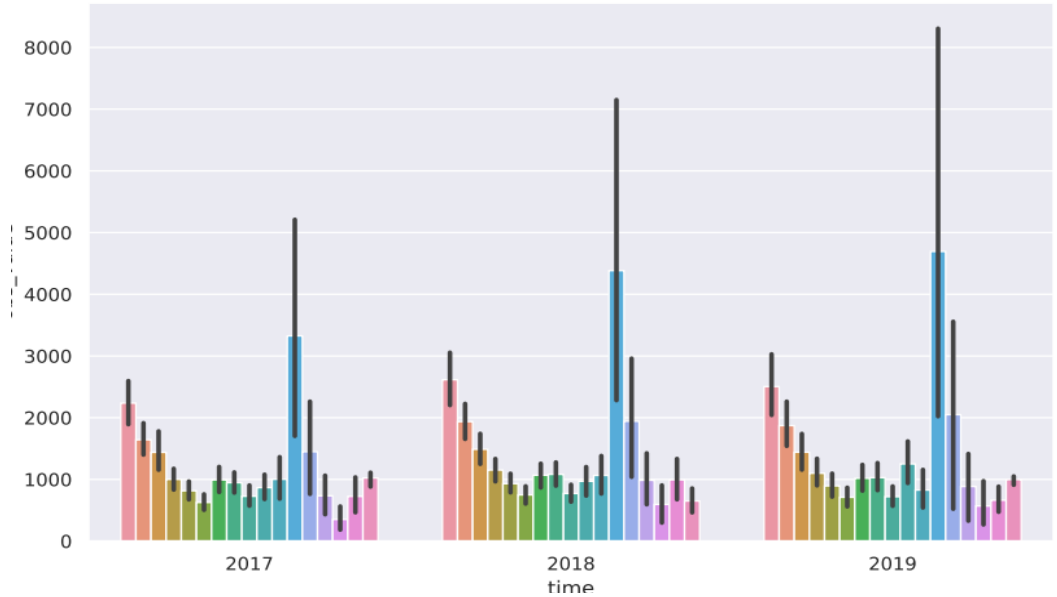
Листинг

```
import numpy as np
import pandas
import matplotlib.pyplot as plt
import seaborn as sns

data =
pandas.read_csv('EAR_4MTH_SEX_OCU_CUR_NB_A.csv'
)
data = data.loc[data['ref_area'] != 'AUS']

data_classif1 = pandas.read_csv('classif1_en.csv', index_col =
'classif1')
data_classif1 = data_classif1.rename(columns={'
classif1.label':'classif1_label'})
data_classif1 = data_classif1[['classif1_label']]

fdata = data.join(data_classif1, on = ['classif1'])
col = "classif1_label"
data =
data[['obs_value','ref_area','sex','classif2','time','classif1_label'
,'classif1']]
first_data[col] = first_data[col].str.split('.').str[1]
first_data = first_data[~first_data.isin([np.nan]).any(1)]
```

Исключаем из рассмотрения Австралию т.к в прошлой работе было выяснено, что данные по этой стране недостоверны.
Чтобы лучше понимать информацию, отображенную на графике, добавим колонку с названиями сфер деятельности. Удаляем строки с NaN.

Данные по миру. Средняя ежемесячная зарплата за период с 2013 и 2015 год и с 2017 до 2019 в долларах США с разбивкой по роду занятий.





Managers
Professionals
Technicians and associate professionals
Clerical support workers
Service and sales workers
Skilled agricultural, forestry and fishery workers
Craft and related trades workers
Plant and machine operators, and assemblers
Elementary occupations
Armed forces occupations
Not elsewhere classified
Legislators, senior officials and managers
Clerks
Service workers and shop and market sales workers
Skilled agricultural and fishery workers
Plant and machine operators and assemblers
Armed forces

Можно заметить, что в период с 2013 по 2014 год средняя ежемесячная зарплата менеджеров лидировала среди других профессий, однако с 2015 лидировать по зарплатам начинают законодатели, высшие должностные лица и руководители.
По не усреднённым данным видно, что ежемесячная зарплата законодателей в 2013 всё же больше. В 2014 ситуация обратная, лидируют менеджеры, а позже разрыв увеличивается и заметен резкий рост зарплат законодателей.
Профессионалы и технические специалисты стабильно остаются на 2 месте по средний ежемесячным зарплатам, однако с 2017 года их догоняют клерки(конторщики, офисные работники).

1

```
data = data[(data.classif2 == 'CUR_TYPE_USD') &
(data.time > 2010) & (data.time < 2014)]
sns.set()
sns.barplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data
);
plt.legend(loc='upper left', bbox_to_anchor=(0, -0.2),
          shadow=True, ncol=2,borderaxespad=0.)
plt.savefig('1.pdf')

data = data[(data.classif2 == 'CUR_TYPE_USD') &
(data.time > 2016) & (data.time < 2020)]
sns.set()
sns.barplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data
);
plt.legend(loc='upper left', bbox_to_anchor=(0, -0.2),
          shadow=True, ncol=2,borderaxespad=0.)
plt.savefig('2.pdf')
```

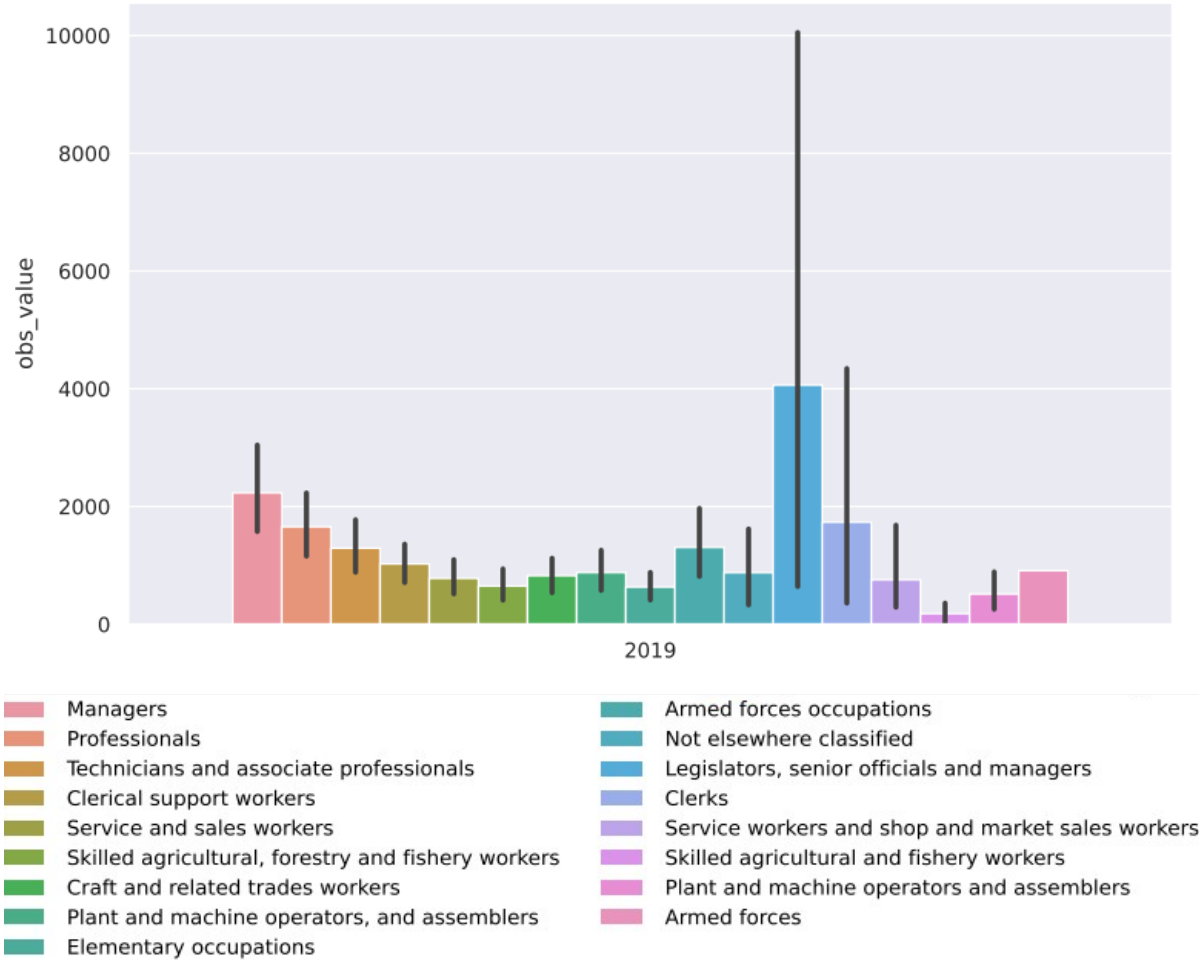
Рассмотрим подробнее как менялась *ежемесячная заработная плата трёх самых высокооплачиваемых специальностей на 2019 год.*

Можно заметить резкий рост ежемесячных зарплат данных сфер деятельности в 2000 году, а так же резкий рост зарплаты менеджеров после 2015 и в 2020.

2

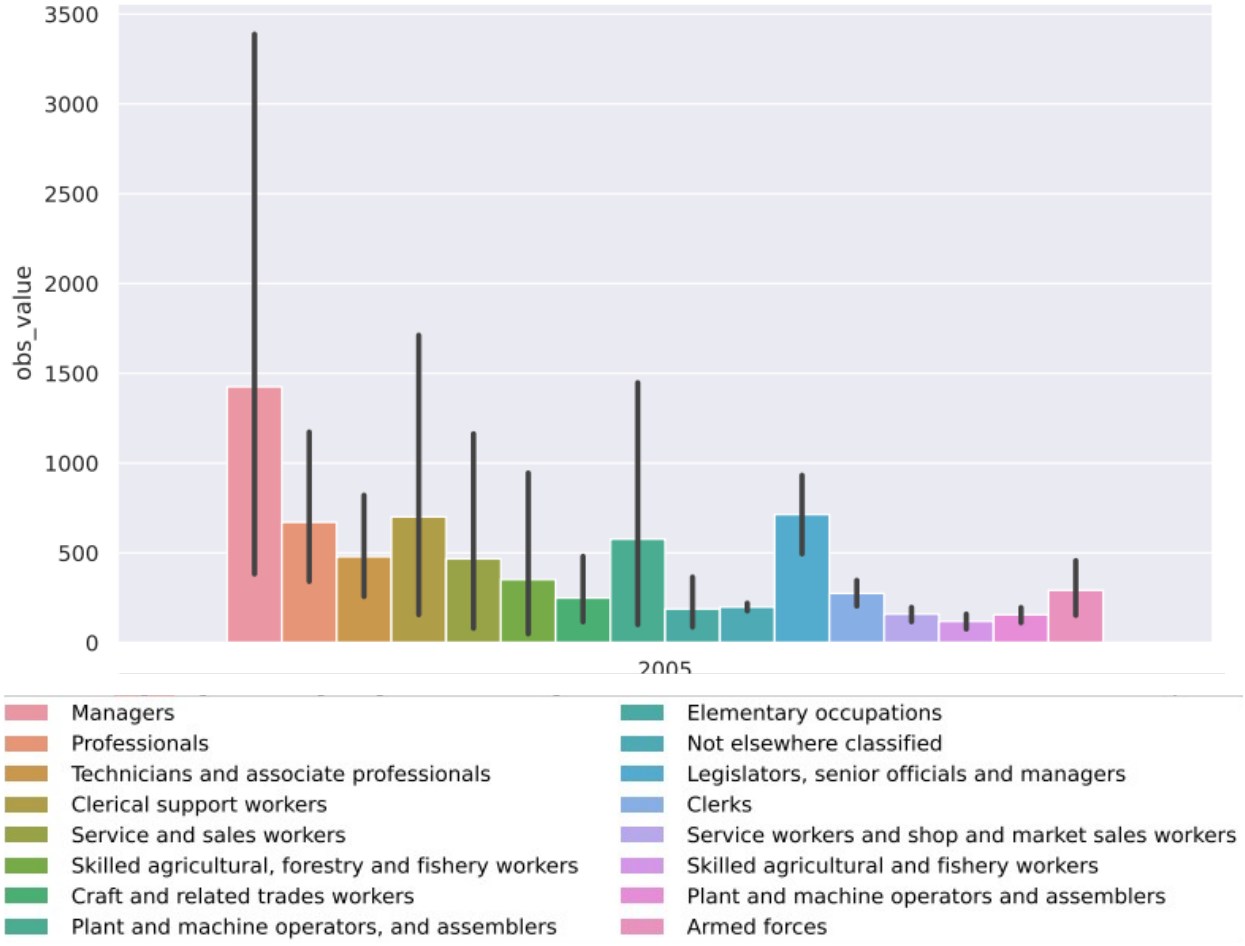
```
data = data[(data.classif2 == 'CUR_TYPE_USD') &
((data.classif1 == 'OCU_ISCO08_1') | (data.classif1 ==
'OCU_ISCO88_4')| (data.classif1 == 'OCU_ISCO88_1'))]
sns.set()
f = sns.relplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data,
    kind = 'line',
    ci = None
);
plt.savefig('5.pdf')
```

Средняя ежемесячная зарплата женщин в 2019 году в долларах США с разбивкой по профессиональной деятельности за 2019 год.



По данному графику можно отметить, что размер зарплат законодателей сильно разниться, но по среднему значению превосходит остальные. Аналогичная ситуация с зарплатами клерков.
Если смотреть по минимальному значению зарплат, то лидирует зарплата менеджеров.

Средняя ежемесячная зарплата женщин в 2019 году в долларах США с разбивкой по профессиональной деятельности за 2005 год.



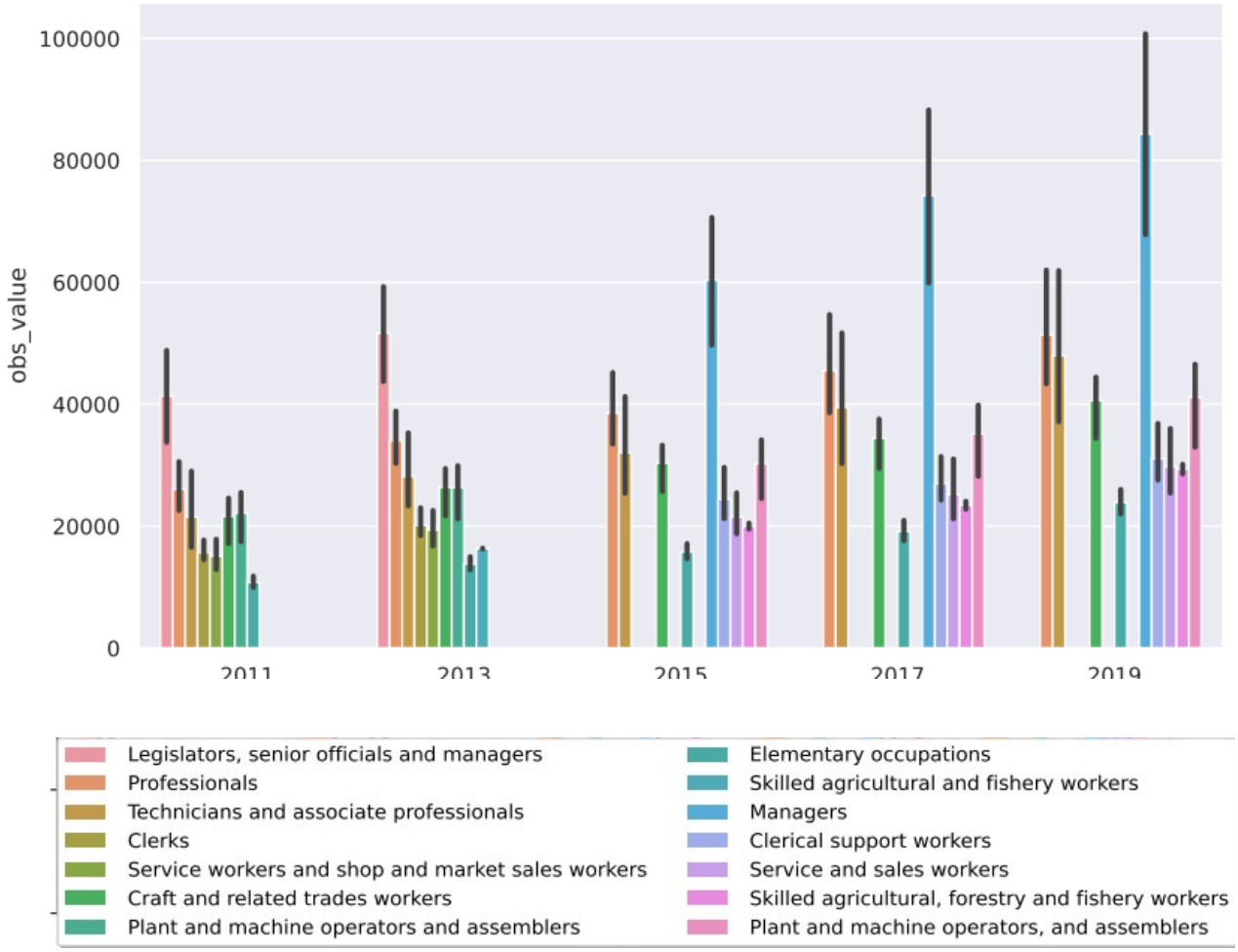
Если сравнивать два графика выше можно заметить, что для женщин ежемесячная зарплата с учётом распределения по сферам деятельности схожа с данными по миру за те же периоды времени.
В 2005 можно заметить , что средние ежемесячные зарплаты законодателей, высших должностных лиц и руководителей и офисных работников примерно равны. Но не усреднённые зарплаты офисных работников всё же больше.
Виден огромный скачок зарплат в сторону законодателей, высших должностных лиц в по сравнению с 2005, где лидируют менеджеры.

3

```
data= data[(data.classif2 == 'CUR_TYPE_USD') &
(data.time == 2019) & (data.sex == 'SEX_F')]
sns.set()
sns.barplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data,
);
plt.legend(loc='upper left', bbox_to_anchor=(0, -0.2),
          shadow=True, ncol=2,borderaxespad=0.)
plt.savefig('3.pdf')
```

```
data= data[(data.classif2 == 'CUR_TYPE_USD') &
(data.time == 2005) & (data.sex == 'SEX_F')]
sns.set()
sns.barplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data,
);
plt.legend(loc='upper left', bbox_to_anchor=(0, -0.2),
          shadow=True, ncol=2,borderaxespad=0.)
plt.savefig('4.pdf')
```

Данные по России в рублях за период от 2010 до 2020 с разбивкой по сфере деятельность.

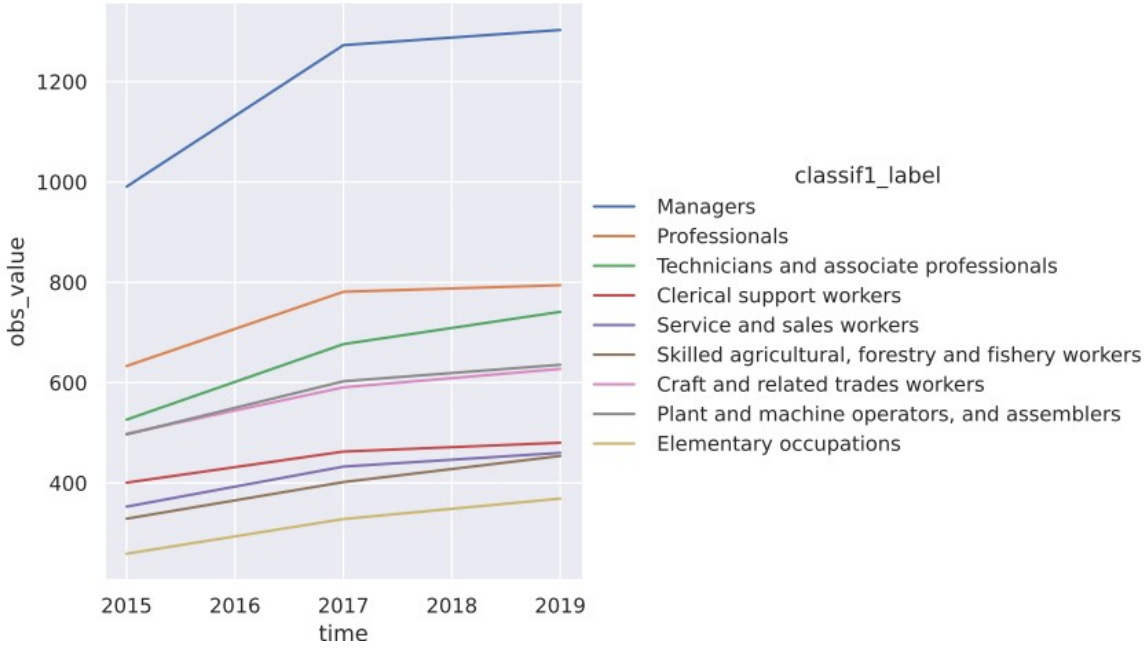


По России сохраняются те же тенденции, что и по миру. Из отличительных особенностей можно отметить отрыв по зарплатам в 2015-2019 профессионалов,технических специалистов и клерков по сравнению с другими специальностями. Ещё можно отметить, что по сравнению с данными по миру зарплаты операторов машин и оборудования располагаются на 3 месте в то время как в мире зарплаты этой специальности располагаются ближе к концу списка(подразумевается список, который мы глядя на график можем для себя составить по величине зарплат по сферам деятельности).

4

```
data_sex_ocu = first_data[(first_data.classif2 ==
'CUR_TYPE_LCU') & (first_data.time > 2010) &
(first_data.ref_area == 'RUS')]
plt.figure(figsize=(10,6))
sns.set()

sns.barplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=data_sex_ocu,
);
plt.legend(loc='upper left', bbox_to_anchor=(0, -0.05),
           shadow=True, ncol=2,borderaxespad=0.)
plt.savefig('6.pdf')
```

Изменение зарплат за период 2015 по 2019 год в России в рублях.



По данному графику можно оценить насколько резко росла зарплата у той или иной специальности.

5

```
first_data = first_data[(first_data.classif2 ==
'CUR_TYPE_USD') & ((first_data.classif1 ==
'OCU_ISCO08_1') | (first_data.classif1 ==
'OCU_ISCO88_4')| (first_data.classif1 ==
'OCU_ISCO88_1'))]
sns.set()
f = sns.relplot(
    x="time",
    y="obs_value",
    hue="classif1_label",
    data=first_data,
    kind = 'line',
    ci = None
);
f.fig.autofmt_xdate()
plt.savefig('5.pdf')
```