

## Знакомство с языком R: анализ данных

1. Установите R (<https://www.r-project.org/>). Для удобства можете поставить среду разработки RStudio.

2. Установите библиотеку Swirl, команда

```
install.packages("swirl")
```

3. Запустите обучение, и следуйте подсказкам:

```
library("swirl")
swirl()
```

Помощник предложит вам пройти на выбор курсы. Курсы доступны в репозитории <http://swirlstats.com/scn/>. Пройдите ознакомительный курс "R Programming" (первый в выборе).

4. После прохождения курса, загрузите в память csv-датасет "titanic\_train.csv". В файле приведены данные пассажиров Титаника, описание столбцов:

- **PassengerId** - ИД пассажира (порядковый номер)
- **Survived** - "0" - если пассажир не пережил крушение корабля
- **Pclass** - "1","2","3" - класс размещения пассажира
- **Name** - имя пассажира
- **Sex** - пол пассажира
- **Age** - возраст
- **SibSp** - количество родных на борту (братьев / сестер / супругов)
- **Parch** - Количество родных на борту (родителей / детей)
- **Ticket** - Номер билета
- **Fare** - стоимость билета
- **Cabin** - номер каюты
- **Embarked** - порт, в котором пассажир произвел посадку. C - Cherbourg, S - Southampton, Q = Queenstown

5. Ответьте на вопросы ниже. В каждом пункте предлагается построить картинку и с ее помощью ответить на вопрос в форме. Конечно, можно попытаться ответить на все вопросы только с обработкой данных, без картинок, но мы советуем Вам потренироваться строить (красивые) визуализации.

1. Сколько мужчин / женщин находилось на борту?
2. Выведите распределение переменной Pclass (социально-экономический статус) и это же распределение, только для мужчин / женщин по отдельности. Сколько было мужчин 2-го класса?
3. Каковы медиана и стандартное отклонение платежей (Fare)? Округлите до 2 десятичных знаков.
4. Правда ли, что люди моложе 30 лет выживали чаще, чем люди старше 60 лет? Каковы доли выживших в обеих группах?
5. Правда ли, что женщины выживали чаще мужчин? Каковы доли выживших в обеих группах?
6. Найдите самое популярное имя среди пассажиров Титаника мужского пола?
7. Сравните графически распределение стоимости билетов и возраста у спасенных и у погибших. Средний возраст погибших выше, верно? Для корректного отображения распределения сперва разделите информацию о стоимости билетов на группы, например: 0-50\$, 50-100\$, 100-200\$,...
8. Как отличается средний возраст мужчин / женщин в зависимости от класса обслуживания?
9. В каком порту на борт село больше всего женщин по отношению к общему числу пассажиров севших на борт в этом порту?
10. Правда ли что чем больше у пассажира родных на борту тем больше вероятность выжить? Проверьте эту теорию с обоими параметрами **Parch** и **SibSp**.