# Анализ статистических данных в MATLAB

Проанализируем данные выборки людей, полученные из соцсети VK

**Содержание**



## Загрузка данных

Загружаем из файла `friends.mat` информацию о людях и их друзьях

```
load friends
```

Чтобы получить свои данные воспользуйтесь скриптом `get_friends`:

```
open get_friends
```

В загруженных данных:

- `source` - ID пользователя-источника, из друзей которого получен данный пользователь
- `id` - уникальный идентификатор (ID) пользователя
- `sex` - пол (1 - Ж, 2 - М)
- `birthday` - дата рождения
- `city` - город проживания
- `country` - страна проживания
- `status` - текстовый статус пользователя

## Фильтрация данных

(используются в т.ч. функции из *Statistics and Machine Learning Toolbox*)

Получим количество друзей для каждого пользователя-"источника"

```
C = groupsummary(data, 'source')
```

C = 168×2 table

| | source | GroupCount |
|---|---|---|
| 1 | 60 | 499 |
| 2 | 411 | 522 |
| 3 | 540 | 985 |
| 4 | 1165 | 230 |

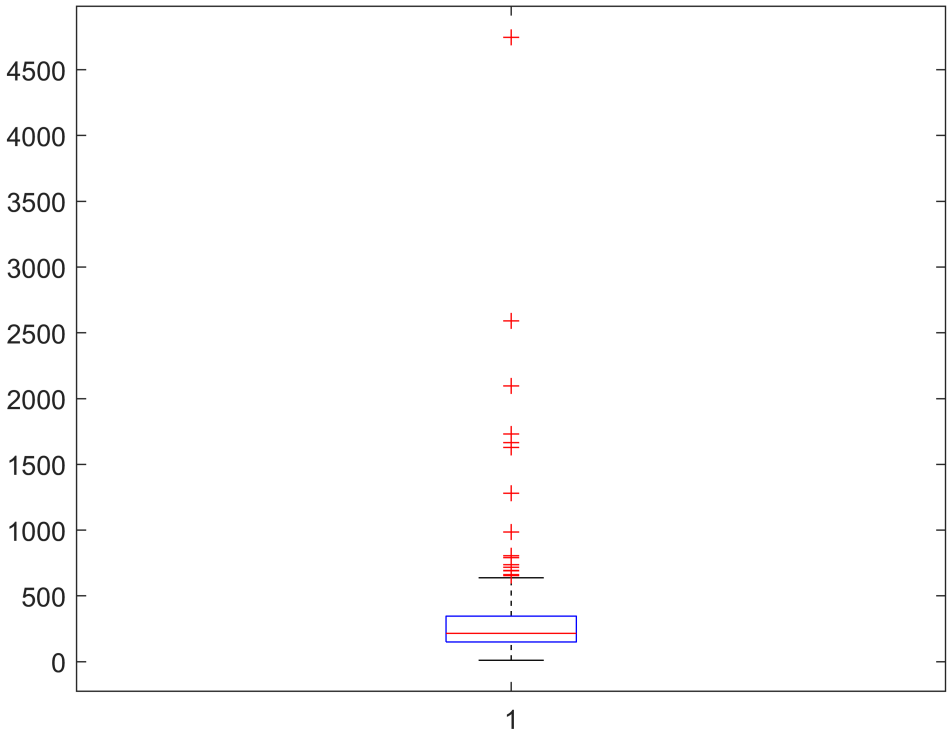
| | source | GroupCount |
|---|---|---|
| 5 | 1385 | 792 |
| 6 | 1676 | 260 |
| 7 | 1756 | 154 |
| 8 | 1856 | 230 |
| 9 | 2106 | 354 |
| 10 | 2444 | 332 |
| 11 | 2673 | 663 |
| 12 | 2996 | 282 |
| 13 | 3073 | 182 |
| 14 | 3083 | 265 |
| 15 | 3173 | 230 |
| 16 | 3333 | 694 |
| 17 | 3382 | 294 |
| 18 | 3387 | 525 |
| 19 | 3495 | 150 |
| 20 | 3516 | 444 |
| 21 | 3906 | 496 |
| 22 | 3959 | 242 |
| 23 | 4396 | 179 |
| 24 | 4528 | 253 |
| 25 | 4678 | 59 |
| 26 | 4975 | 176 |
| 27 | 5177 | 370 |
| 28 | 5247 | 169 |
| 29 | 5351 | 100 |
| 30 | 5353 | 182 |
| 31 | 5490 | 557 |
| 32 | 5607 | 124 |
| 33 | 5646 | 268 |
| 34 | 5957 | 113 |
| 35 | 6046 | 429 |
| 36 | 6305 | 690 |
| 37 | 6600 | 98 |
| 38 | 6605 | 115 |

| | source | GroupCount |
|---|---|---|
| 39 | 6607 | 260 |
| 40 | 6846 | 216 |
| 41 | 6972 | 242 |
| 42 | 6988 | 249 |
| 43 | 6994 | 197 |
| 44 | 7116 | 488 |
| 45 | 7354 | 2590 |
| 46 | 7418 | 250 |
| 47 | 7535 | 454 |
| 48 | 7568 | 190 |
| 49 | 7589 | 325 |
| 50 | 7607 | 224 |
| 51 | 7667 | 278 |
| 52 | 7675 | 164 |
| 53 | 7738 | 383 |
| 54 | 7771 | 201 |
| 55 | 7906 | 267 |
| 56 | 7980 | 312 |
| 57 | 8193 | 500 |
| 58 | 8645 | 188 |
| 59 | 8647 | 196 |
| 60 | 8654 | 177 |
| 61 | 8856 | 197 |
| 62 | 9167 | 352 |
| 63 | 9311 | 138 |
| 64 | 9478 | 195 |
| 65 | 9666 | 277 |
| 66 | 9703 | 195 |
| 67 | 9753 | 139 |
| 68 | 10377 | 737 |
| 69 | 10554 | 11 |
| 70 | 10827 | 1630 |
| 71 | 10945 | 320 |
| 72 | 11023 | 183 |

| | source | GroupCount |
|---|---|---|
| 73 | 11126 | 175 |
| 74 | 11300 | 134 |
| 75 | 11624 | 392 |
| 76 | 12003 | 186 |
| 77 | 12036 | 85 |
| 78 | 12170 | 165 |
| 79 | 12472 | 805 |
| 80 | 12658 | 399 |
| 81 | 12722 | 206 |
| 82 | 12730 | 386 |
| 83 | 12841 | 192 |
| 84 | 13286 | 57 |
| 85 | 13294 | 158 |
| 86 | 13654 | 142 |
| 87 | 13796 | 341 |
| 88 | 13937 | 140 |
| 89 | 14185 | 143 |
| 90 | 14896 | 58 |
| 91 | 15014 | 275 |
| 92 | 15096 | 145 |
| 93 | 15331 | 455 |
| 94 | 15475 | 631 |
| 95 | 15949 | 264 |
| 96 | 16314 | 225 |
| 97 | 16457 | 187 |
| 98 | 16571 | 217 |
| 99 | 17203 | 297 |
| 100 | 17445 | 179 |

⋮

```
count = C.GroupCount;
boxplot(count);
```




Найдем верхнюю статистическую границу количества друзей

```
upper = iqr(count) * 1.5 + quantile(count, 0.75)
```

```
upper = 641.2500
```

Находим "источники", у которых количество друзей превышает границу (выбросы)

```
bad_sources = C.source(count > upper);
```

Удаляем из таблицы всех друзей этих "источников"

```
bad_rows = ismember(data.source, bad_sources);
data(bad_rows, :) = [];
```

Извлекаем из таблицы связи между пользователями

```
relations = data(:, {'source', 'id'});
data.source = [];
```

Удаляем повторяющиеся строки из таблицы

```
data = unique(data, 'rows');
```

Переведем данные о поле (`sex`) в массив категорий (`categorical`)

```
data.sex = categorical(data.sex, [1 2], ["Female", "Male"]);
```

Найдем пользователей, у которых не указан пол

```
no_sex = ismissing(data.sex);
nnz(no_sex)
```

```
ans = 4
```

Удалим из таблицы пользователей, у которых не указан пол

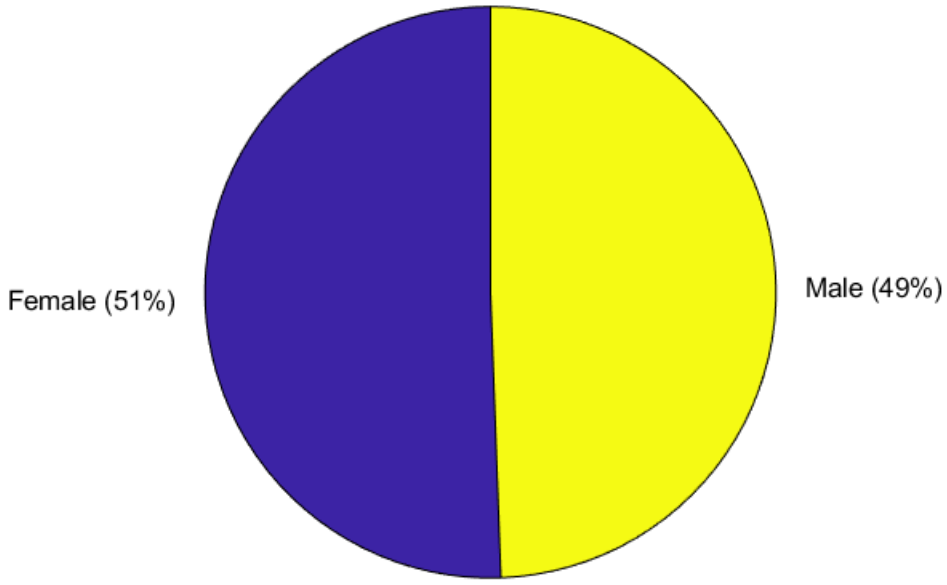
```
data = rmmissing(data, 'DataVariables', 'sex');
```

# Анализ и визуализация

## Анализ пола

Построим круговую диаграмму

```
figure
pie(data.sex)
```




## Анализ возраста

Переведем дату рождения в формат `datetime`

```
data.birthday = datetime(data.birthday);
```

Запишем возраст каждого человека в новый столбец age

```
data.age = years(datetime('today') - data.birthday);
```

Найдем статистику по возрасту в зависимости от пола

```
age_stat = groupsummary(data, 'sex', {'min', 'max', 'median'}, 'age')
```

age_stat = 2×5 table

| | sex | GroupCount | min_age | max_age | median_age |
|---|---|---|---|---|---|
| 1 | Female | 13262 | 14.4206 | 117.3056 | 28.8000 |
| 2 | Male | 12959 | 14.3795 | 117.3440 | 29.5694 |

Классифицируем людей по возрасту (в соответствии с ВОЗ)

```
edges = [0 18 66 80 100 150]
```

edges = 1×6

| 0 | 18 | 66 | 80 | 100 | 150 |
|---|---|---|---|---|---|

```
cats = ["kid", "young", "middle", "senior", "old"]
```

cats = *1×5 string array*

| "kid" | "young" | "middle" | "senior" | "old" |
|---|---|---|---|---|

```
data.agecat = discretize(data.age, edges, 'categorical', cats)
```

data = 26221×9 table

...

| | id | first_name | sex | birthday | city | country | status |
|---|---|---|---|---|---|---|---|
| 1 | 5 | 'Данил' | Male | 18-Oct-1986 | 'Санкт-Пе... | 'Россия' | '"Даже ес... |
| 2 | 6 | 'Марина' | Female | 11-Jun-1989 | 'Москва' | 'Россия' | 'Время-то... |
| 3 | 7 | 'Дмитрий' | Male | 04-May-1988 | 'Санкт-Пе... | 'Россия' | '' |
| 4 | 8 | 'Варвара' | Female | NaT | 'Москва' | 'Россия' | '' |
| 5 | 10 | 'Максим' | Male | 10-Nov-1981 | 'Москва' | 'Россия' | '' |
| 6 | 11 | 'Алексей' | Male | NaT | 'Санкт-Пе... | 'Россия' | 'Консульт... |
| 7 | 15 | 'Андрей' | Male | 25-Sep-1989 | 'Санкт-Пе... | 'Россия' | 'тихо шур... |
| 8 | 16 | 'Гаврилов' | Male | 26-May-1989 | 'Санкт-Пе... | 'Россия' | '' |
| 9 | 18 | 'Евгения' | Female | NaT | '' | '' | '' |
| 10 | 20 | 'Анна' | Female | NaT | 'Москва' | 'Россия' | '' |
| 11 | 21 | 'Антон' | Male | 24-Sep-1988 | 'Новосибирск' | 'Россия' | 'Shit hap... |
| 12 | 23 | 'Michael' | Male | NaT | 'Москва' | 'Россия' | '' |
| 13 | 25 | 'Надюша' | Female | NaT | 'Москва' | 'Россия' | 'Vivo! Vi... |
| 14 | 26 | 'Коля' | Male | NaT | 'Санкт-Пе... | 'Россия' | 'дружба э... |

| | id | first_name | sex | birthday | city | country | status |
|---|---|---|---|---|---|---|---|
| 15 | 28 | 'Александр' | Male | 11-Nov-1986 | '' | '' | '-Fly, fi... |
| 16 | 30 | 'Михаил' | Male | NaT | 'Москва' | 'Россия' | 'Если отп... |
| 17 | 31 | 'Алексей' | Male | NaT | 'Москва' | 'Россия' | '«Не согл... |
| 18 | 32 | 'Irene' | Female | NaT | 'Москва' | 'Россия' | 'и мечты.... |
| 19 | 33 | 'Елена' | Female | NaT | 'Москва' | 'Россия' | '' |
| 20 | 34 | 'Саркис' | Male | NaT | 'Москва' | 'Россия' | '' |
| 21 | 35 | 'Валентина' | Female | NaT | 'Москва' | 'Россия' | 'Умные по... |
| 22 | 36 | 'Кирилл' | Male | NaT | 'Москва' | 'Россия' | '' |
| 23 | 40 | 'Александр' | Male | NaT | 'Москва' | 'Россия' | '' |
| 24 | 41 | 'Evelina' | Female | NaT | '' | '' | '' |
| 25 | 42 | 'Николай' | Male | NaT | 'Москва' | 'Россия' | '' |
| 26 | 44 | 'Вадим' | Male | 17-Aug-1989 | 'Москва' | 'Россия' | 'Кот руки... |
| 27 | 48 | 'Вадим' | Male | 02-Nov-1989 | 'Санкт-Пе... | 'Россия' | '' |
| 28 | 49 | 'Даша' | Female | 18-Jul-1989 | 'Санкт-Пе... | 'Россия' | 'Качестве... |
| 29 | 52 | 'Ксения' | Female | NaT | 'Москва' | 'Россия' | '' |
| 30 | 58 | 'Игорь' | Male | 24-Oct-1986 | 'Москва' | 'Россия' | '' |
| 31 | 60 | 'Владимир' | Male | 29-Dec-1984 | 'Москва' | 'Россия' | 'Кавер-гр... |
| 32 | 62 | 'Михаил' | Male | NaT | 'Москва' | 'Россия' | '' |
| 33 | 64 | 'Михаил' | Male | 26-Aug-1985 | 'Москва' | 'Россия' | '' |
| 34 | 65 | 'Артём' | Male | NaT | 'Москва' | 'Россия' | '#' |
| 35 | 67 | 'Евгений' | Male | 21-Sep-1988 | '' | '' | 'изменить... |
| 36 | 69 | 'Мария' | Female | 15-Feb-1986 | 'Севастополь' | 'Россия' | 'Если воо... |
| 37 | 70 | 'Александра' | Female | NaT | 'Ногинск' | 'Россия' | 'Наша тре... |
| 38 | 71 | 'Дмитрий' | Male | NaT | 'Москва' | 'Россия' | 'друзья и... |
| 39 | 72 | 'Виталий' | Male | 18-Oct-1987 | 'Москва' | 'Россия' | '' |
| 40 | 77 | 'Юлия' | Female | NaT | 'Москва' | 'Россия' | 'Когда че... |
| 41 | 84 | 'Александр' | Male | NaT | 'Москва' | 'Россия' | '' |
| 42 | 94 | 'Алексей' | Male | NaT | 'Москва' | 'Россия' | 'Get_Nature' |
| 43 | 96 | 'Мария' | Female | NaT | 'Санкт-Пе... | 'Россия' | 'Художест... |
| 44 | 99 | 'Дмитрий' | Male | NaT | 'Москва' | 'Россия' | '' |
| 45 | 100 | 'Владимир' | Male | NaT | 'Москва' | 'Россия' | 'Всей Сав... |
| 46 | 101 | 'Александр' | Male | NaT | 'Москва' | 'Россия' | '' |
| 47 | 103 | 'Михаил' | Male | NaT | 'Москва' | 'Россия' | '' |
| 48 | 106 | 'Алексей' | Male | NaT | 'Москва' | 'Россия' | '' |

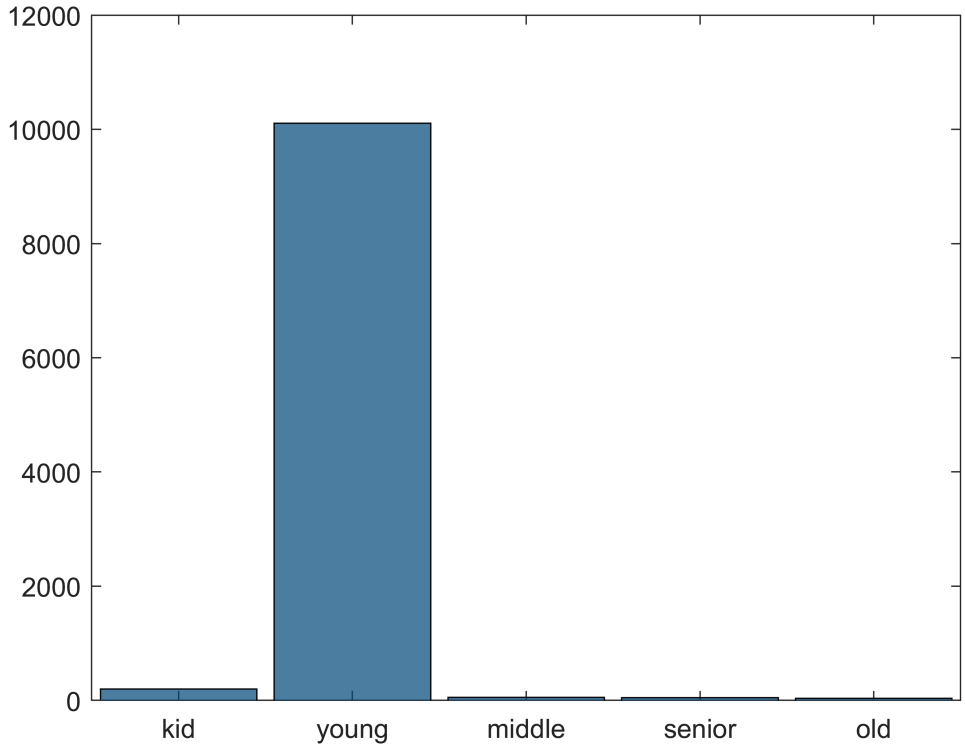
| | id | first_name | sex | birthday | city | country | status |
|---|---|---|---|---|---|---|---|
| 49 | 113 | 'Элеонора' | Female | NaT | 'Москва' | 'Россия' | '' |
| 50 | 115 | 'Алексей' | Male | NaT | 'Москва' | 'Россия' | '' |
| 51 | 116 | 'Иван' | Male | 12-Sep-1986 | 'Санкт-Пе... | 'Россия' | '' |
| 52 | 117 | 'Сергей' | Male | 14-Oct-1982 | 'Москва' | 'Россия' | '' |
| 53 | 120 | 'Дарья' | Female | NaT | 'Москва' | 'Россия' | 'Зорко од... |
| 54 | 123 | 'Алексей' | Male | 28-Dec-1984 | '' | 'Россия' | '' |
| 55 | 125 | 'Юрий' | Male | NaT | 'Москва' | 'Россия' | '' |
| 56 | 131 | 'Павел' | Male | 10-Apr-1984 | 'Москва' | 'Россия' | 'Падение ... |
| 57 | 132 | 'Александр' | Male | 02-Dec-1989 | 'Москва' | 'Россия' | '' |
| 58 | 134 | 'Александра' | Female | 24-Jun-1986 | '' | 'Таиланд' | 'Групповы... |
| 59 | 135 | 'Александр' | Male | NaT | 'Москва' | 'Россия' | '' |
| 60 | 137 | 'Дарья' | Female | 28-Jan-1989 | 'Paris' | 'Франция' | '' |
| 61 | 139 | 'Алексей' | Male | NaT | 'Москва' | 'Россия' | '' |
| 62 | 141 | 'Ирина' | Female | 03-Jan-1989 | 'Москва' | 'Россия' | 'Lose you... |
| 63 | 143 | 'Даша' | Female | NaT | '' | '' | 'Выходя и... |
| 64 | 145 | 'Ярослав' | Male | 06-Sep-1989 | 'Москва' | 'Россия' | '' |
| 65 | 146 | 'Светлана' | Female | NaT | 'Москва' | 'Россия' | '' |
| 66 | 148 | 'Diman' | Male | NaT | 'Москва' | 'Россия' | '' |
| 67 | 149 | 'Игорь' | Male | NaT | 'Москва' | 'Россия' | 'У вас че... |
| 68 | 151 | 'Наима' | Male | NaT | 'New York... | 'США' | 'продолжа... |
| 69 | 154 | 'Yulia' | Female | NaT | 'Петах Тиква' | 'Израиль' | 'Jerusale... |
| 70 | 156 | 'Татьяна' | Female | 05-Feb-1990 | 'Москва' | 'Россия' | '' |
| 71 | 158 | 'Елена' | Female | NaT | '' | '' | '' |
| 72 | 159 | 'Федосэ' | Male | 11-Sep-1989 | 'Москва' | 'Россия' | '' |
| 73 | 161 | 'Вадим' | Male | 27-Mar-1988 | 'Москва' | 'Россия' | 'Всем про... |
| 74 | 162 | 'Наталия' | Female | 02-Jun-1987 | 'Bergen' | 'Норвегия' | '' |
| 75 | 164 | 'Птица' | Female | NaT | 'Москва' | 'Россия' | 'https://... |
| 76 | 168 | 'Vikus' | Female | 06-Sep-1989 | 'Москва' | 'Россия' | '' |
| 77 | 169 | 'Денис' | Male | 05-May-1978 | 'Москва' | 'Россия' | 'Если вы ... |
| 78 | 175 | 'Игорь' | Male | NaT | 'Москва' | 'Россия' | 'Релаксир... |
| 79 | 178 | 'Кирилл' | Male | 03-May-1989 | 'Zürich' | 'Швейцария' | '' |
| 80 | 179 | 'Люся' | Female | 28-Jun-1989 | 'Москва' | 'Россия' | 'А если к... |
| 81 | 182 | 'Илья' | Male | NaT | 'Москва' | 'Россия' | 'спасибо,... |
| 82 | 184 | 'Александр' | Male | NaT | 'Москва' | 'Россия' | 'fb.com/p... |

| | id | first_name | sex | birthday | city | country | status |
|---|---|---|---|---|---|---|---|
| 83 | 185 | 'Екатерина' | Female | 11-Jun-1989 | 'Москва' | 'Россия' | '' |
| 84 | 187 | 'Павел' | Male | NaT | '' | '' | '' |
| 85 | 188 | 'Тимофей' | Male | 16-Sep-1988 | 'Москва' | 'Россия' | '' |
| 86 | 190 | 'Василий' | Male | 27-Sep-1987 | 'Санкт-Пе... | 'Россия' | 'В мире т... |
| 87 | 191 | 'Екатерина' | Female | 12-Jun-1987 | 'Москва' | 'Россия' | 'Фотограф... |
| 88 | 195 | 'Андрей' | Male | NaT | 'Москва' | 'Россия' | 'Well, th... |
| 89 | 196 | 'Василий' | Male | 18-Aug-1985 | 'Москва' | 'Россия' | '' |
| 90 | 197 | 'Ия' | Female | NaT | '' | 'Россия' | 'Свидетел... |
| 91 | 199 | 'Яков' | Male | NaT | 'Москва' | 'Россия' | '' |
| 92 | 200 | 'Стю' | Female | NaT | 'Москва' | 'Россия' | '' |
| 93 | 202 | 'Евгения' | Female | NaT | 'Москва' | 'Россия' | 'Сообщени... |
| 94 | 204 | 'Денис' | Male | NaT | 'Москва' | 'Россия' | '' |
| 95 | 205 | 'Евгения' | Female | NaT | 'Москва' | 'Россия' | '' |
| 96 | 207 | 'Антон' | Male | 05-Jun-1989 | '' | 'Россия' | '' |
| 97 | 208 | 'Игорь' | Male | NaT | 'Москва' | 'Россия' | '¡Carpe d... |
| 98 | 210 | 'Ирина' | Female | 27-Aug-1987 | 'Москва' | 'Россия' | 'Looking ... |
| 99 | 211 | 'Ирина' | Female | 22-Feb-1990 | '' | 'Россия' | 'Джеронимо!' |
| 100 | 213 | 'Антоха' | Male | NaT | 'Москва' | 'Россия' | 'Knowledg... |

⋮

Строим гистограмму по категориям

```
figure
histogram(data.agecat)
```

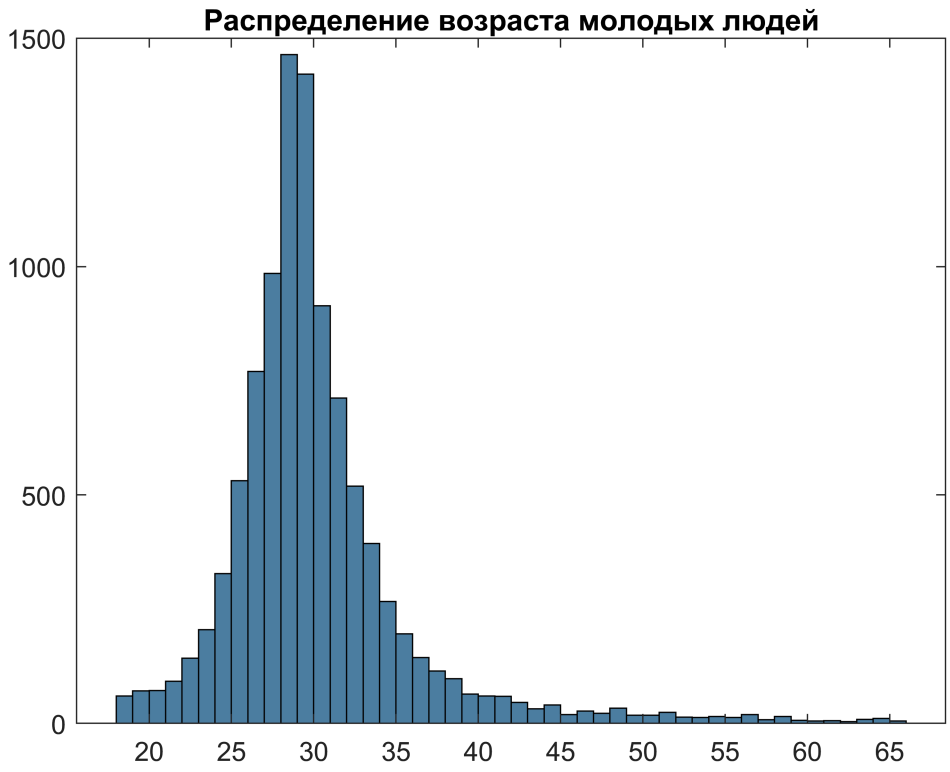



Вытащим из данных молодых людей (young)

```
young = data(data.agecat == "young", 'age');
```

Построим гистограмму возраста для молодых

```
figure
histogram(young.age)
title("Распределение возраста молодых людей")
```




## Исследование связей

Подготовим данные для построения графа друзей

```
EdgeTable = mergevars(relations, {'id', 'source'}, 'NewVariableName', 'EndNodes')
```

EdgeTable = 35202×1 table

| | EndNodes | |
|---|---|---|
| 1 | 60 | 3073 |
| 2 | 411 | 3073 |
| 3 | 540 | 3073 |
| 4 | 1165 | 3073 |
| 5 | 1385 | 3073 |
| 6 | 1676 | 3073 |
| 7 | 1756 | 3073 |
| 8 | 1856 | 3073 |
| 9 | 2106 | 3073 |
| 10 | 2444 | 3073 |
| 11 | 2673 | 3073 |
| 12 | 2996 | 3073 |

| | EndNodes | |
|---|---|---|
| 13 | 3083 | 3073 |
| 14 | 3173 | 3073 |
| 15 | 3333 | 3073 |
| 16 | 3382 | 3073 |
| 17 | 3387 | 3073 |
| 18 | 3495 | 3073 |
| 19 | 3516 | 3073 |
| 20 | 3906 | 3073 |
| 21 | 3959 | 3073 |
| 22 | 4396 | 3073 |
| 23 | 4528 | 3073 |
| 24 | 4678 | 3073 |
| 25 | 4918 | 3073 |
| 26 | 4971 | 3073 |
| 27 | 4975 | 3073 |
| 28 | 5177 | 3073 |
| 29 | 5247 | 3073 |
| 30 | 5351 | 3073 |
| 31 | 5353 | 3073 |
| 32 | 5490 | 3073 |
| 33 | 5607 | 3073 |
| 34 | 5646 | 3073 |
| 35 | 5648 | 3073 |
| 36 | 5957 | 3073 |
| 37 | 6046 | 3073 |
| 38 | 6305 | 3073 |
| 39 | 6600 | 3073 |
| 40 | 6605 | 3073 |
| 41 | 6607 | 3073 |
| 42 | 6846 | 3073 |
| 43 | 6972 | 3073 |
| 44 | 6988 | 3073 |
| 45 | 6994 | 3073 |
| 46 | 7116 | 3073 |

| | EndNodes | |
|---|---|---|
| 47 | 7354 | 3073 |
| 48 | 7418 | 3073 |
| 49 | 7535 | 3073 |
| 50 | 7568 | 3073 |
| 51 | 7589 | 3073 |
| 52 | 7607 | 3073 |
| 53 | 7667 | 3073 |
| 54 | 7675 | 3073 |
| 55 | 7738 | 3073 |
| 56 | 7771 | 3073 |
| 57 | 7906 | 3073 |
| 58 | 7980 | 3073 |
| 59 | 8193 | 3073 |
| 60 | 8645 | 3073 |
| 61 | 8647 | 3073 |
| 62 | 8654 | 3073 |
| 63 | 8856 | 3073 |
| 64 | 9167 | 3073 |
| 65 | 9311 | 3073 |
| 66 | 9478 | 3073 |
| 67 | 9666 | 3073 |
| 68 | 9703 | 3073 |
| 69 | 9753 | 3073 |
| 70 | 9859 | 3073 |
| 71 | 10377 | 3073 |
| 72 | 10554 | 3073 |
| 73 | 10827 | 3073 |
| 74 | 10945 | 3073 |
| 75 | 11023 | 3073 |
| 76 | 11126 | 3073 |
| 77 | 11300 | 3073 |
| 78 | 11624 | 3073 |
| 79 | 12003 | 3073 |
| 80 | 12036 | 3073 |

| | EndNodes | |
|---|---|---|
| 81 | 12170 | 3073 |
| 82 | 12472 | 3073 |
| 83 | 12658 | 3073 |
| 84 | 12722 | 3073 |
| 85 | 12730 | 3073 |
| 86 | 12841 | 3073 |
| 87 | 13286 | 3073 |
| 88 | 13294 | 3073 |
| 89 | 13654 | 3073 |
| 90 | 13796 | 3073 |
| 91 | 13937 | 3073 |
| 92 | 14185 | 3073 |
| 93 | 14896 | 3073 |
| 94 | 14915 | 3073 |
| 95 | 15014 | 3073 |
| 96 | 15096 | 3073 |
| 97 | 15331 | 3073 |
| 98 | 15475 | 3073 |
| 99 | 15949 | 3073 |
| 100 | 16314 | 3073 |
| ⋮ | | |

Создаем граф

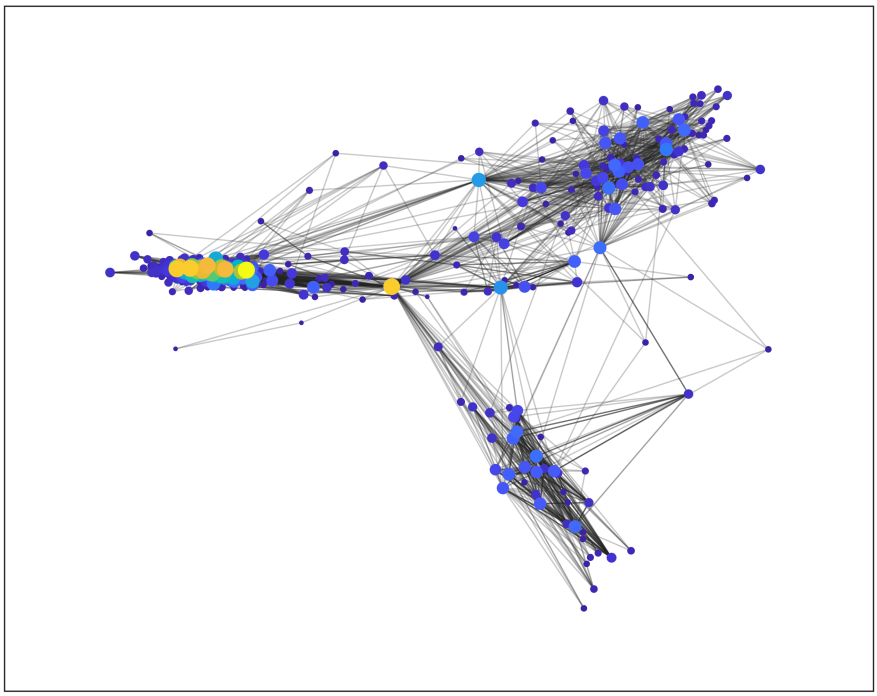
```
G = graph(EdgeTable);
```

Упрощаем граф

```
G = simplify(G);
G = rmnode(G, find(degree(G) < 4));
```

Визуализируем граф

```
figure
g = plot(G, 'EdgeColor', [.7 .7 .7], 'EdgeAlpha', 0.2);
g.MarkerSize = log(degree(G)) + 0.1;
g.NodeCData = degree(G);
```



## Анализ местоположения

В таблице создадим новый столбец с местоположением пользователя

```
data = convertvars(data, {'city', 'country'}, 'string');
data = standardizeMissing(data, "", 'DataVariables', {'city', 'country'});
data.location = data.country + ", " + data.city;
```

Загружаем географичесие координаты (геокоды) пользователей

```
load geodata.mat
```

Если вы используете свои данные пользователей, необходимо получить для них свои геокоды:

```
open get_locations
```

Определим, сколько человек живет в каждой локации

```
C0 = groupsummary(data, 'location');
```

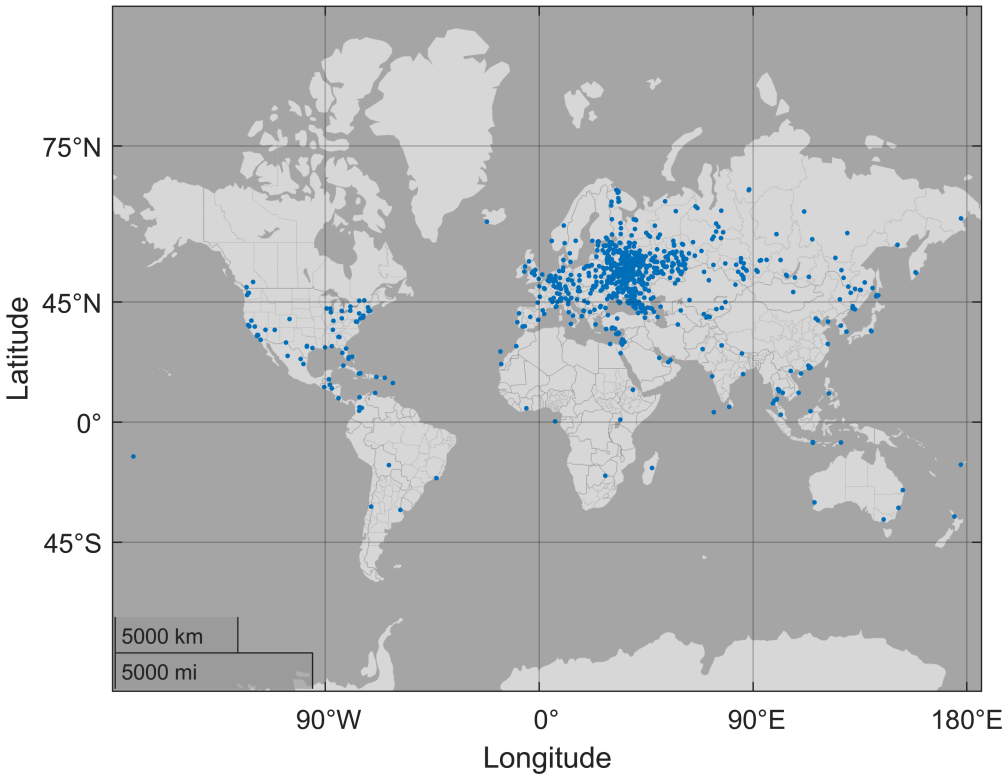
Объединим эти данные с таблицей геокодов

```
C = innerjoin(C0, geodata);
```

Отобразим все локации на карте мира

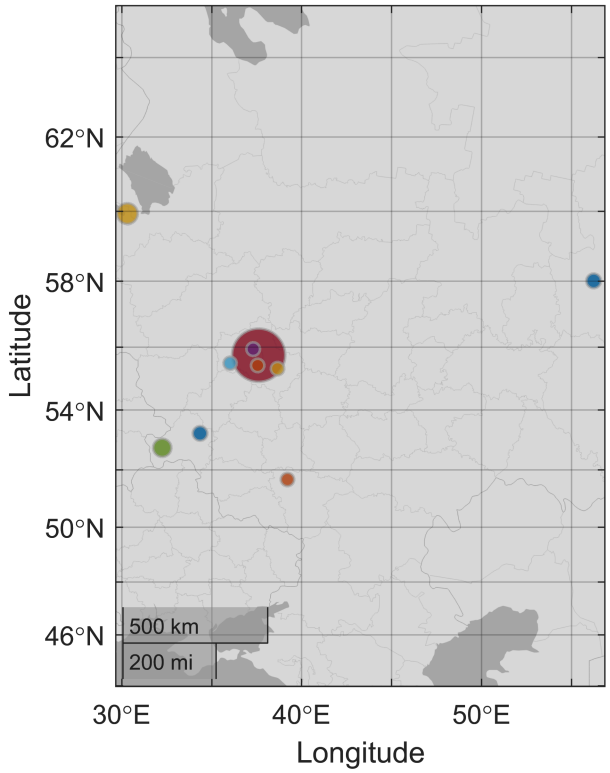
```
figure
```

```
geoscatter(C.lat, C.long, '.')
```




Отобразим топ 10 городов по количеству человек

```
figure
C10 = topkrows(C, 10, 'GroupCount');
C10.location = categorical(C10.location);
geobubble(C10, 'lat', 'long', 'SizeVariable', 'GroupCount', 'ColorVariable', 'location');
legend
```




## Текстовый анализ статусов

(*Text Analytics Toolbox*)

Некоторые пользователи ставят в соцсети текстовые статусы.

Объединим их в один текст

```
s = string(data.status);
s = join(s);
```

Предобработаем текст

```
s = lower(s);
s = strrep(s, 'ё', 'е');
s = erasePunctuation(s);
```

Токенизируем текст и проведем обработку

```
doc = tokenizedDocument(s);
```

```
Warning: No supported language detected (based on script or alphabet), setting the language to 'en'. For tips
 showing how to use Text Analytics with languages other than English, German, and Japanese, see Language
 Considerations.
```

```
doc = removeLongWords(doc, 15);
doc = removeShortWords(doc, 2);
doc = removeWords(doc, ["чтобы"]);
```

Составляем из токенов "мешок" фраз по 3 слова

```
bn = bagOfNgrams(doc, 'NgramLengths', 3);
```

Визуализируем "облако слов"

```
figure
wc = wordcloud(bn);
```