

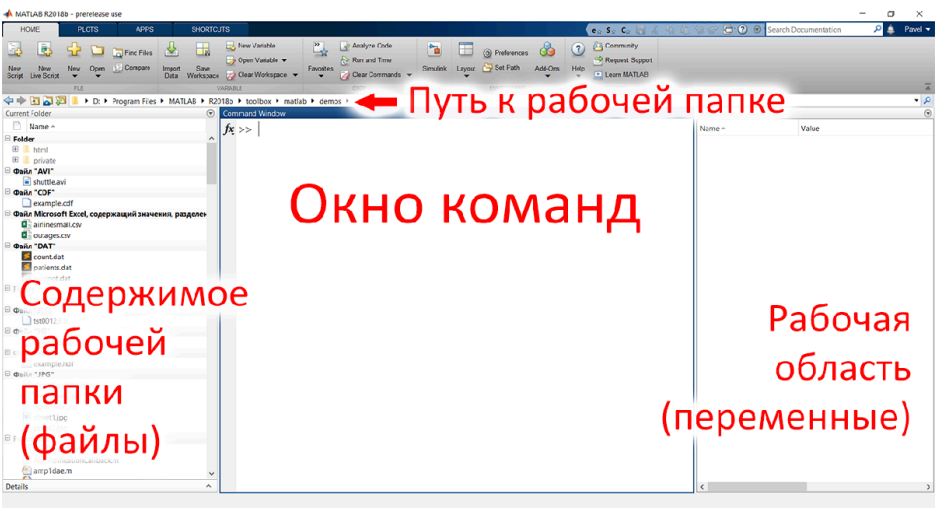
Анализ данных в MATLAB для начинающих

Проанализируем статистику цунами из Excel-файла в MATLAB без использования команд

Содержание

Главное окно MATLAB.....	1
Импорт данных из файла Excel.....	1
Поиск файла.....	1
Мастер импорта данных.....	2
Анализ данных.....	3
Работа с переменной.....	3
Построение графиков.....	5
Сохранение результатов работы.....	8

Главное окно MATLAB



Центральную область занимает **окно команд**, сюда вы будете писать свои команды и тут же получать результат. Слева находится содержимое **текущей рабочей папки**, путь к ней задается в строке над окном команд (**путь к рабочей папке**). Это та папка, в которой для удобства лежат ваши рабочие файлы и ваши MATLAB-программы. В **рабочую область** будут выводиться все переменные MATLAB, с которыми вы работаете.

Импорт данных из файла Excel

Поиск файла

Статистика цунами записана в файле `tsunamis.xlsx`, который идет вместе с MATLAB для примера и лежит в папке его установки.

Откроем эту папку командой

```
winopen(fullfile(matlabroot, 'toolbox\matlab\demos'))
```

Скопируем её путь, и вставим в строку **пути к рабочей папке** в главном окне, чтобы сделать папку рабочей и упростить доступ к файлу.

Либо можно сменить путь командой

```
cd(fullfile(matlabroot, 'toolbox\matlab\demos'))
```

Находим в окне **содержимого рабочей папки** файл `tsunamis.xlsx`, и кликаем по нему 2 раза, чтобы загрузить, откроется окно мастера импорта.

Также открыть окно импорта нашего файла можно командой

```
uiimport('tsunamis.xlsx')
```

Мастер импорта данных

The screenshot shows the MATLAB Import Wizard interface. The 'IMPORT' tab is active. The 'Range' is set to 'A2:T163'. The 'Output Type' is set to 'Table'. The 'Replace' checkbox is checked. The 'unimportable cells with' dropdown is set to 'NaN'. The 'Import Selection' button is highlighted with a green checkmark. The 'Variable Names Row' is set to '1'. The 'Text Options' button is visible. The 'UNIMPORTABLE CELLS' section is empty. The 'tsunamis.xlsx' file is selected. The 'Выбор листа' (Select Sheet) button is highlighted. The 'Выбранная область импорта' (Selected Import Area) is highlighted. The 'Правила импорта пропусков' (Import Missing Value Rules) section is visible. The 'Генерация кода' (Generate Code) button is visible. The 'Настройка столбцов' (Column Settings) section is visible. The 'Пропуски' (Missing Values) section is visible. The 'Выбор листа' (Select Sheet) button is highlighted.

	Latitude	Longitude	Year	Month	Day	Hour	Minute	Second	ValidityCode	Validity	Cause
1	Latitude	Longitude	Year	Month	Day	Hour	Minute	Second	ValidityCode	Validity	Cause
2	-3.8000	128.3000	1950	10	8	3	23		2	questionabl...	
3									4	definite tsu...	
4	-9.0200	157.9500	1951	12	22				2	questionabl...	
5	42.1500	143.8500	1952	3					4	definite tsu...	
6	19.1000	-155	1952	3					4	definite tsu...	
7	43.1000	-82.4000	1952	5	6				1	very doubtf...	
8	52.7500	159.5000	1952	11	4	16	58		4	definite tsu...	
9	50	156.5000	1953	3	18				3	probable ts...	
10	-2.4000	147.4000	1953	6	27				3	probable ts...	
11	-18.3000	178.2000	1953	9	14	0	26	36	4	definite tsu...	
12	34	141.7000	1953	11	25	17	48		4	definite tsu...	

По виду он чем-то напоминает Excel, вы можете изучить данные и настроить параметры импорта. Например, вы можете вручную указать, какую часть таблицы хотите импортировать. По умолчанию она будет импортирована в переменную типа `table` - это самый удобный формат в MATLAB для хранения разнородных массивов данных. А у нас разные столбики имеют разный тип - число, категория, это может быть текст, дата или другой тип. Тип столбика можно поменять здесь же. Как и тип итоговой

переменной. В таблице есть пропущенные значения, подсвеченные желтым, по-умолчанию они будут заменены на значение NaN (Not a Number).

После настройки жмем зеленую кнопку и получаем переменную `tsunamis`, с которой будем работать. Обратите внимание, что под кнопкой импорта есть дополнительные функции. Мы можем сгенерировать MATLAB скрипт или функцию, которая автоматически сделает все то, что мы настроили мышкой и выдаст такую же переменную. Таким образом мы можем свою работу автоматизировать и применить полученную функцию к другим файлам с такой же структурой.

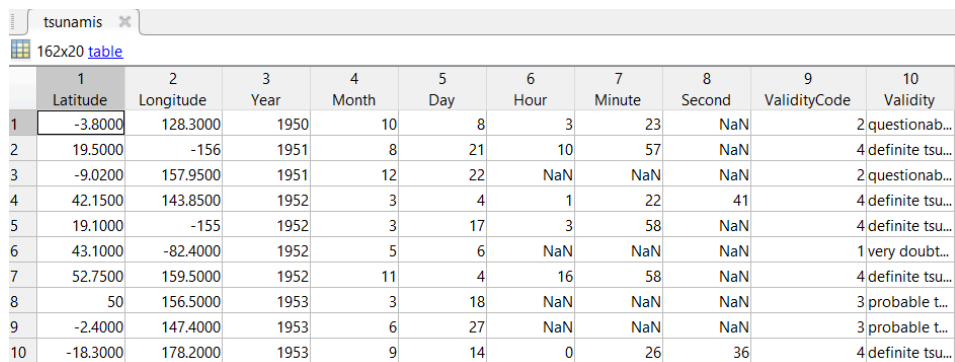
Анализ данных

Работа с переменной

После импорта в **рабочей области** главного окна появилась переменная `tsunamis`. Откроем её двойным кликом и изучим таблицу, которая в ней хранится.

Также переменную можно открыть командой

```
open('tsunamis')
```



	1	2	3	4	5	6	7	8	9	10
	Latitude	Longitude	Year	Month	Day	Hour	Minute	Second	ValidityCode	Validity
1	-3.8000	128.3000	1950	10	8	3	23	NaN		2 questionab...
2	19.5000	-156	1951	8	21	10	57	NaN		4 definite tsu...
3	-9.0200	157.9500	1951	12	22	NaN	NaN	NaN		2 questionab...
4	42.1500	143.8500	1952	3	4	1	22	41		4 definite tsu...
5	19.1000	-155	1952	3	17	3	58	NaN		4 definite tsu...
6	43.1000	-82.4000	1952	5	6	NaN	NaN	NaN		1 very doubt...
7	52.7500	159.5000	1952	11	4	16	58	NaN		4 definite tsu...
8	50	156.5000	1953	3	18	NaN	NaN	NaN		3 probable t...
9	-2.4000	147.4000	1953	6	27	NaN	NaN	NaN		3 probable t...
10	-18.3000	178.2000	1953	9	14	0	26	36		4 definite tsu...

Видим что переменная имеет тип данных `table`, т.е. таблица, 162 строки, т.е. измерения, и 20 столбцов, т.е. параметров цунами. Каждый столбик имеет свое название, считанное из исходной таблицы.

С таблицей можно работать с помощью мышки и клавиатуры. Например, можно изменять значения в ячейках (заменять пропуски). Можно переименовать или поменять местами столбцы, удалить лишние, вставить новые. Так же можно вставлять и удалять строки. По любому столбцу можно сортировать всю таблицу.

Обратите внимание, что все манипуляции с таблицей отражаются в виде кода в окне команд. Таким образом MATLAB подталкивает нас к освоению команд. А ещё мы можем копировать эти команды и использовать в алгоритмах.

Кстати, считать таблицу можно командой:

```
tsunamis = readtable('tsunamis.xlsx')
```

Например, после сортировки по столбцу Year прописывается команда

```
tsunamis = sortrows(tsunamis, 'Year', 'ascend')
```

tsunamis = 162×19 table

	Latitude	Longitude	Year	Month	Day	Hour
1	-3.8000	128.3000	1950	10	8	3
2	19.5000	-156.0000	1951	8	21	10
3	-9.0200	157.9500	1951	12	22	NaN
4	42.1500	143.8500	1952	3	4	1
5	19.1000	-155.0000	1952	3	17	3
6	43.1000	-82.4000	1952	5	6	NaN
7	52.7500	159.5000	1952	11	4	16
8	50.0000	156.5000	1953	3	18	NaN
9	-2.4000	147.4000	1953	6	27	NaN
10	-18.3000	178.2000	1953	9	14	0

⋮

После удаления столбца Second прописывается

```
tsunamis = removevars(tsunamis, 'Second')
```

tsunamis = 162×18 table

	Latitude	Longitude	Year	Month	Day	Hour
1	-3.8000	128.3000	1950	10	8	3
2	19.5000	-156.0000	1951	8	21	10
3	-9.0200	157.9500	1951	12	22	NaN
4	42.1500	143.8500	1952	3	4	1
5	19.1000	-155.0000	1952	3	17	3
6	43.1000	-82.4000	1952	5	6	NaN
7	52.7500	159.5000	1952	11	4	16
8	50.0000	156.5000	1953	3	18	NaN
9	-2.4000	147.4000	1953	6	27	NaN
10	-18.3000	178.2000	1953	9	14	0

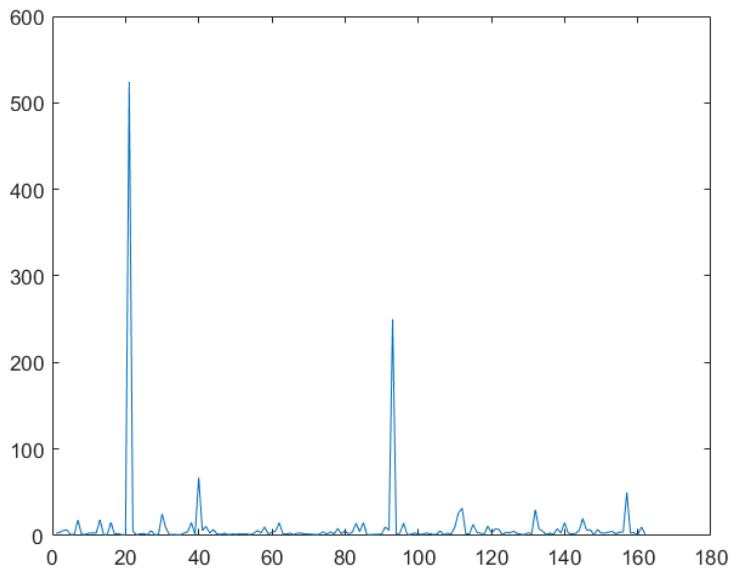
⋮

Построение графиков

В переменной `tsunamis` выделим столбец `MaxHeight`, кликнув на его заголовок. Затем в главном окне MATLAB на вкладке **PLOTS** кликаем на график **plot**. Получаем простой график, где по оси абсцисс отложены порядковые номера точек, по оси ординат - значения столбца `MaxHeight`.

При этом прописывается команда

```
plot(tsunamis.MaxHeight)
```



Теперь выделим столбец `Year` и с зажатой клавишей `Ctrl` выделим также `MaxHeight`. Для двух выделенных столбцов на вкладке **PLOTS** доступен график **scatter**, построим его.

При этом прописывается команда

```
scatter(tsunamis.Year,tsunamis.MaxHeight)
```

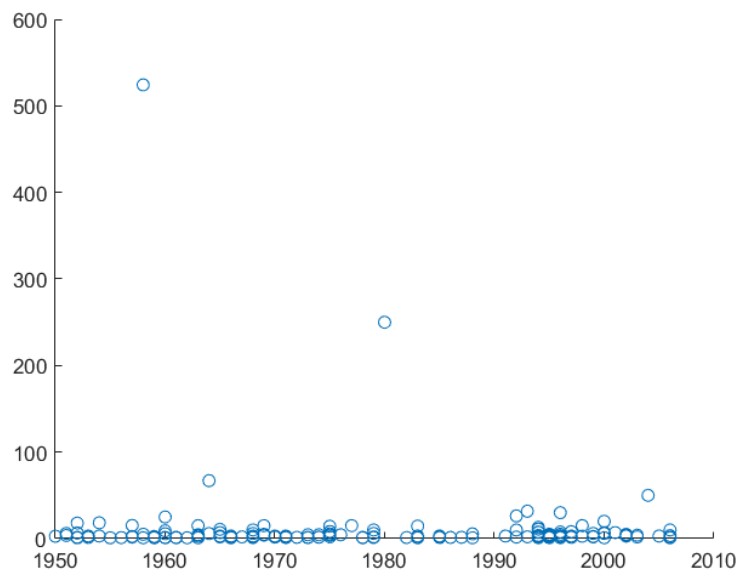


График можно настроить, например, выбрав *Insert -> Title*, подписать его.

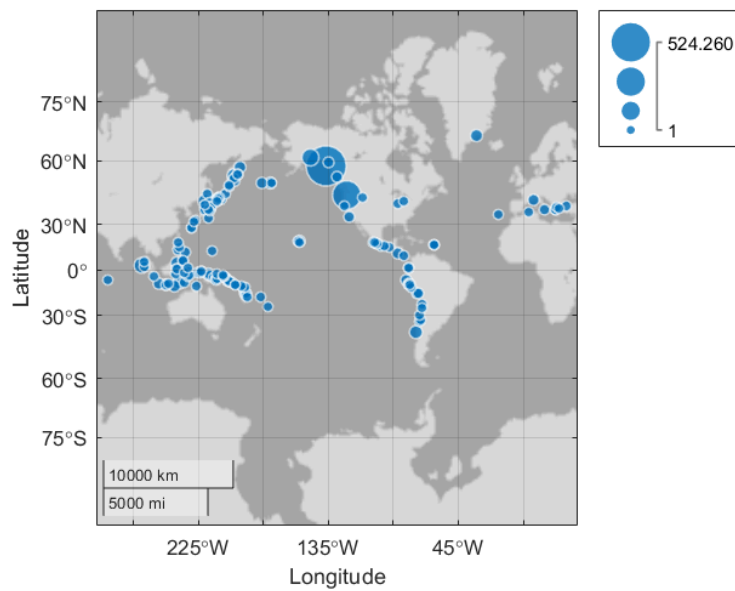
В меню *File -> Save As...* график можно сохранить на диск, чтобы потом, например, вставить в отчет.

Кроме того, *выбрав File -> Generate Code...*, вы получите MATLAB-код, который строит в точности такой же график со всеми настройками и оформлением.

Теперь выделим 3 столбца (с зажатым Ctrl): Latitude, Longitude, MaxHeight. Строим график **geobubble** (на вкладке **PLOTS**).

При этом прописывается команда

```
geobubble(tsunamis.Latitude,tsunamis.Longitude,tsunamis.MaxHeight);
```



Карта позволяет проанализировать, где цунами случаются чаще всего, и где они самые высокие.

Выделим столбец **Country** и построим график **wordcloud**.

При этом прописывается команда

```
wordcloud(tsunamis.Country);
```

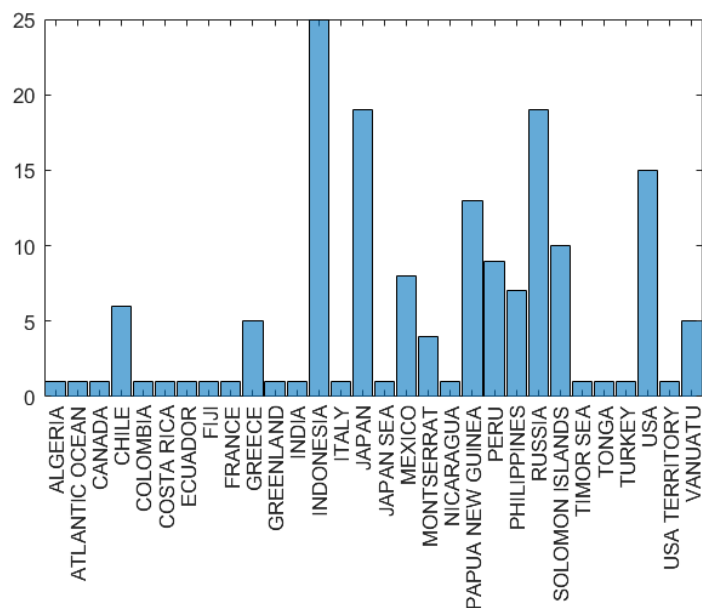


Облако слов наглядно показывает, в каких странах цунами наблюдаются чаще всего.

Оценим количественно, построив гистограмму - график **histogram**.

При этом прописывается команда

```
histogram(tsunamis.Country)
```



В России с 1950 по 2006 года было зафиксировано 19 цунами.

Сохранение результатов работы

Если закрыть MATLAB, переменная `tsunamis` удалится. Чтобы сохранить её на диск можно на вкладке **HOME** нажать кнопку **Save Workspace**, при этом все переменные из **рабочей области** будут сохранены в файл с расширением `.mat`.

Сохранить все переменные в файл `matlab.mat` в текущую рабочую папку можно также командой

```
save
```

Saving to: D:\MATLAB\Data Analysis\COURSE\2. Анализ данных в MATLAB для начинающих\matlab.mat

Чтобы загрузить все данные из `.mat`-файла, надо на него 2 раза кликнуть в окне **содержимого рабочей папки**.

Загрузить данные из файла `matlab.mat` можно также командой

```
load
```


Pavel Roslovets, [ETMC Exponenta](#) © 2018