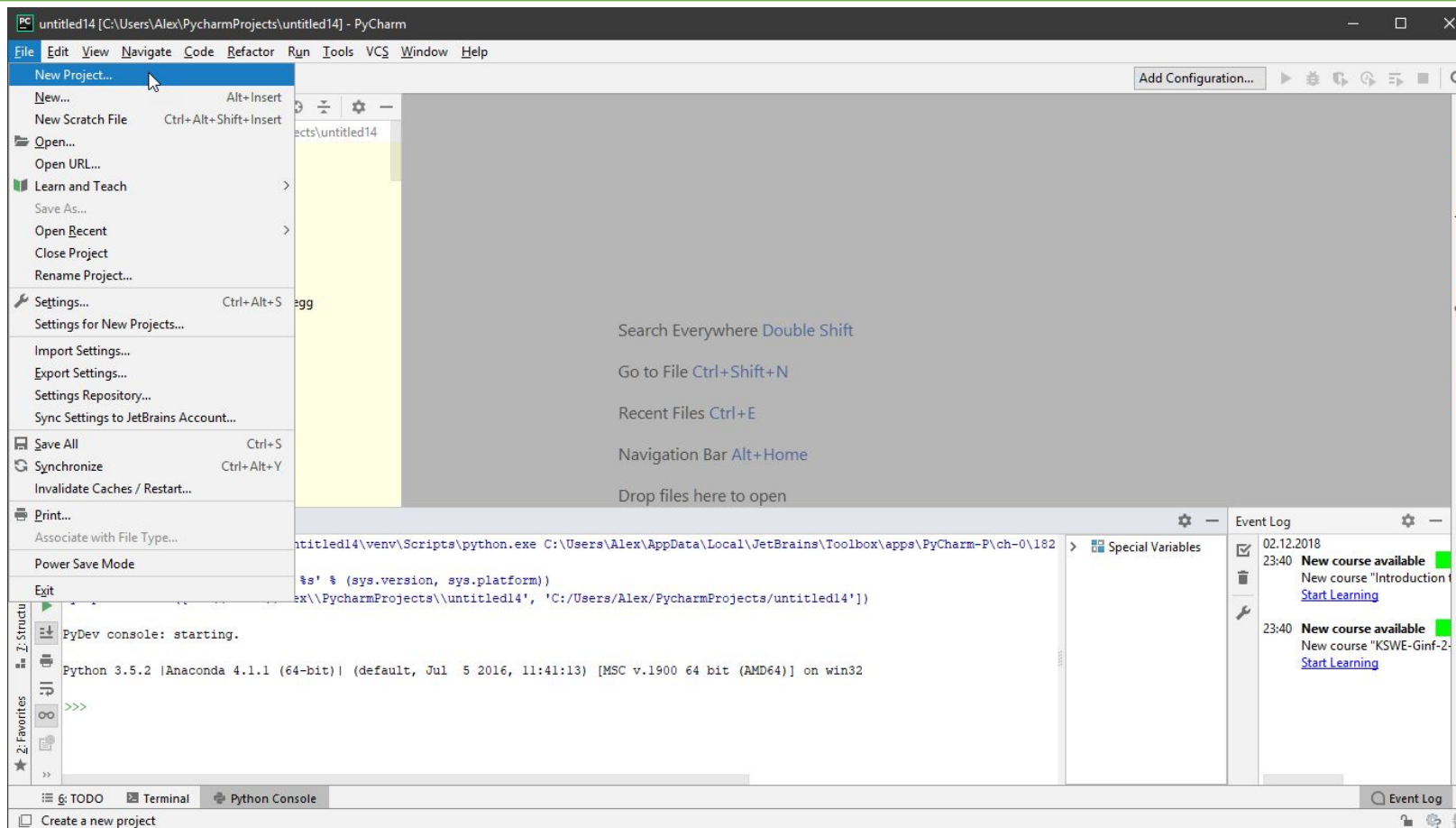
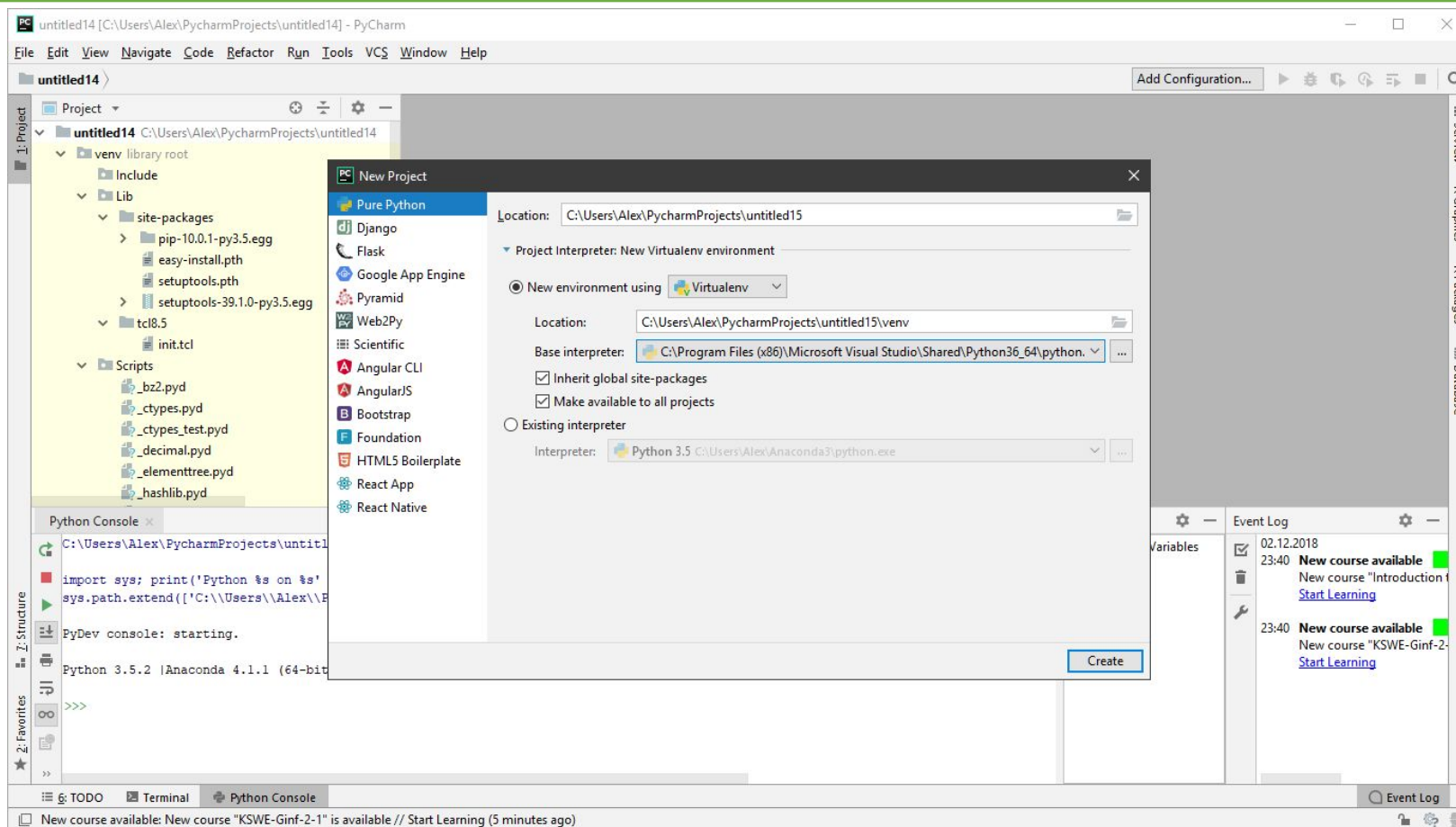


Новый проект



Подключаем или создаем среду разработки python

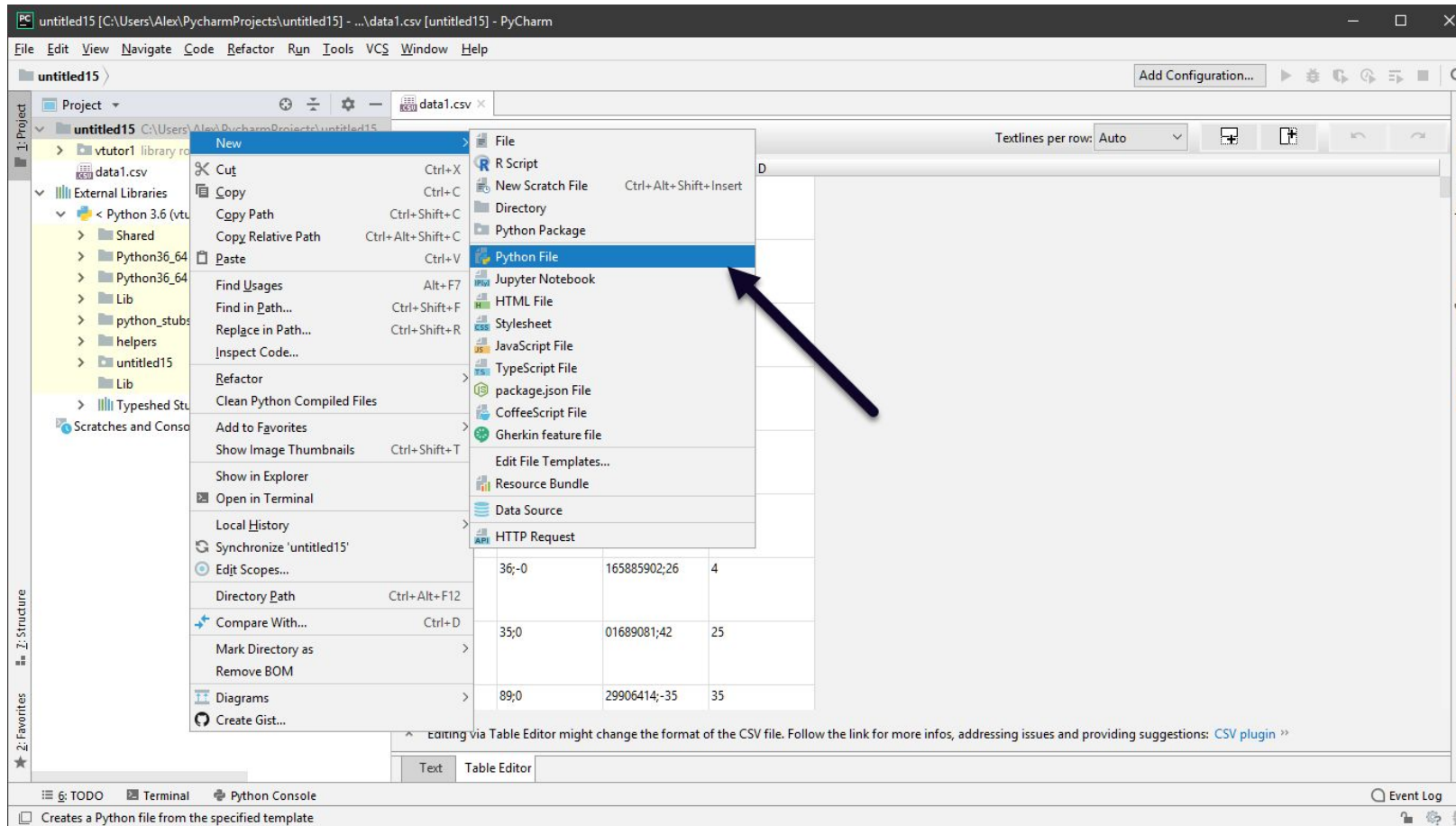


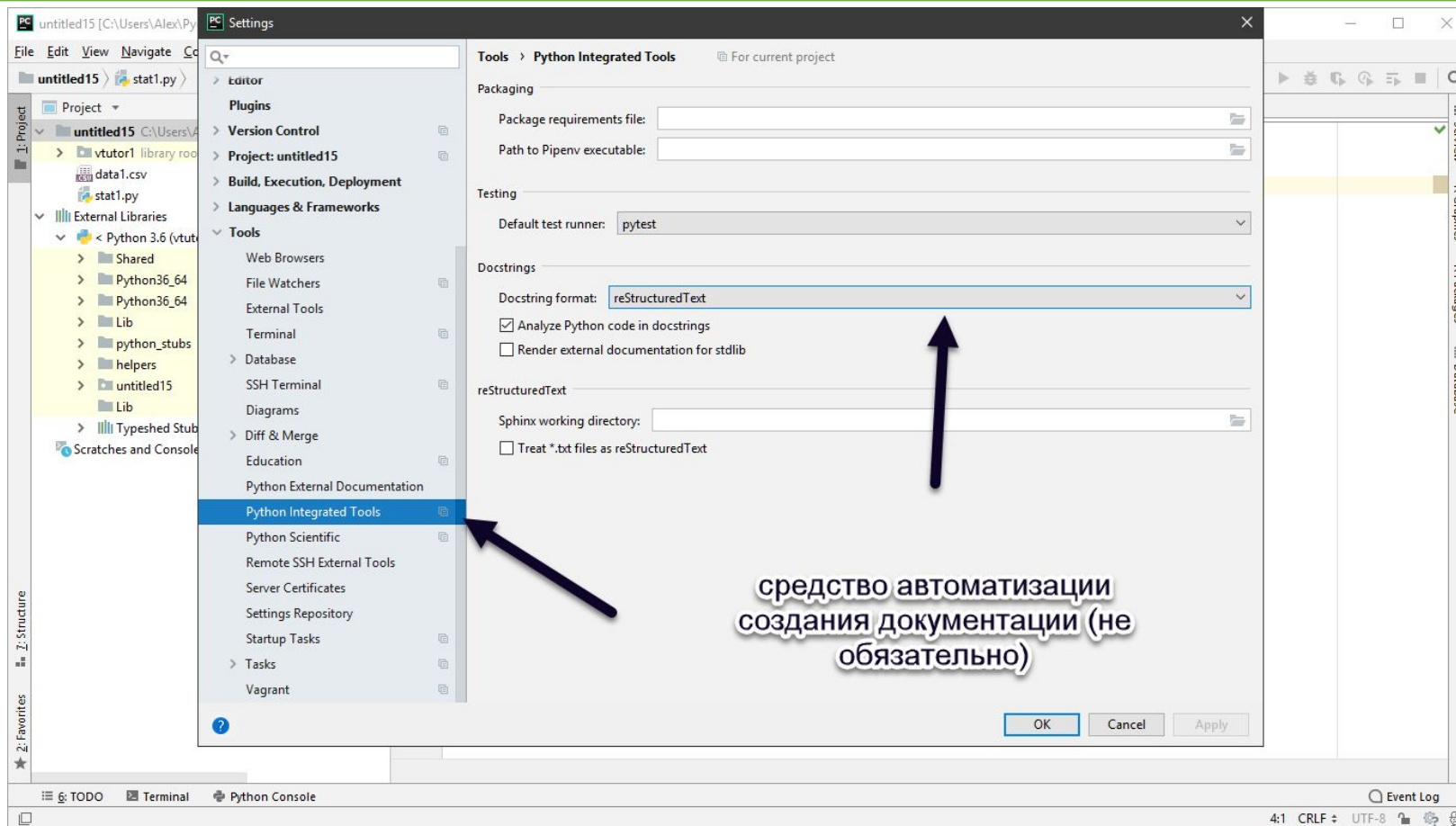
Старый файл с данными

The screenshot shows the PyCharm IDE interface. The main window displays a CSV file named 'data1.csv' in a 'Table Editor' view. The table has four columns labeled A, B, C, and D, and eight rows of data. A red circle '1' points to the 'data1.csv' file in the project tree on the left. A red circle '2' points to the 'Table Editor' tab at the bottom of the editor area. A red circle '3' points to the 'Table Editor' view itself. A notification at the bottom of the editor area reads: 'Editing via Table Editor might change the format of the CSV file. Follow the link for more infos, addressing issues and providing suggestions: [CSV plugin](#)'.

	A	B	C	D
1	1;-2	21;-0	289241179;-32	15
2	2;-0	35;-0	076083492;-3	25
3	3;-1	8;0	054204459;-24	
4	4;0	4;0	007001044;10	
5	5;-2	59;-0	040952326;-33	85
6	6;1	36;-0	165885902;26	4
7	7;2	35;0	01689081;42	25
8	8;-2	89;0	29906414;-35	35

Новый файл python





Внешние модули

- Модули надо установить в текущее окружение
- Подключить к программе
- Среда PyCharm предлагает СНАЧАЛА подключить модуль
- Затем, если его нет в окружении – скачать через Интернет и установить
- После скачивания среда тратит время на анализ модуля, это требует времени

The screenshot displays the PyCharm IDE interface. The main editor window shows a Python file named `stat1.py` with the following content:

```
1 #!/usr/bin/env python
2 #-*- coding: utf-8 -*-
3
4 """
5 Расчет статистики с помощью специальных модулей
6 """
7
8 import numpy as np
```

A tooltip is visible over the `numpy` import, displaying the error message: "No module named numpy more... (Ctrl+F1)".

The left sidebar shows the Project view with the following structure:

- untitled15 (C:\Users\Alex\PycharmProjects\untitled15)
 - vtutor1 (library root)
 - data1.csv
 - stat1.py
 - External Libraries
 - Python 3.6 (vtutor1) > C:\Users\Alex\PycharmProj...
 - Shared
 - Python36_64
 - Python36_64
 - Lib
 - python_stubs
 - helpers
 - untitled15
 - Lib
 - Typeshed Stubs
 - Scratches and Consoles

The bottom panel contains the Python Console, which shows the following output:

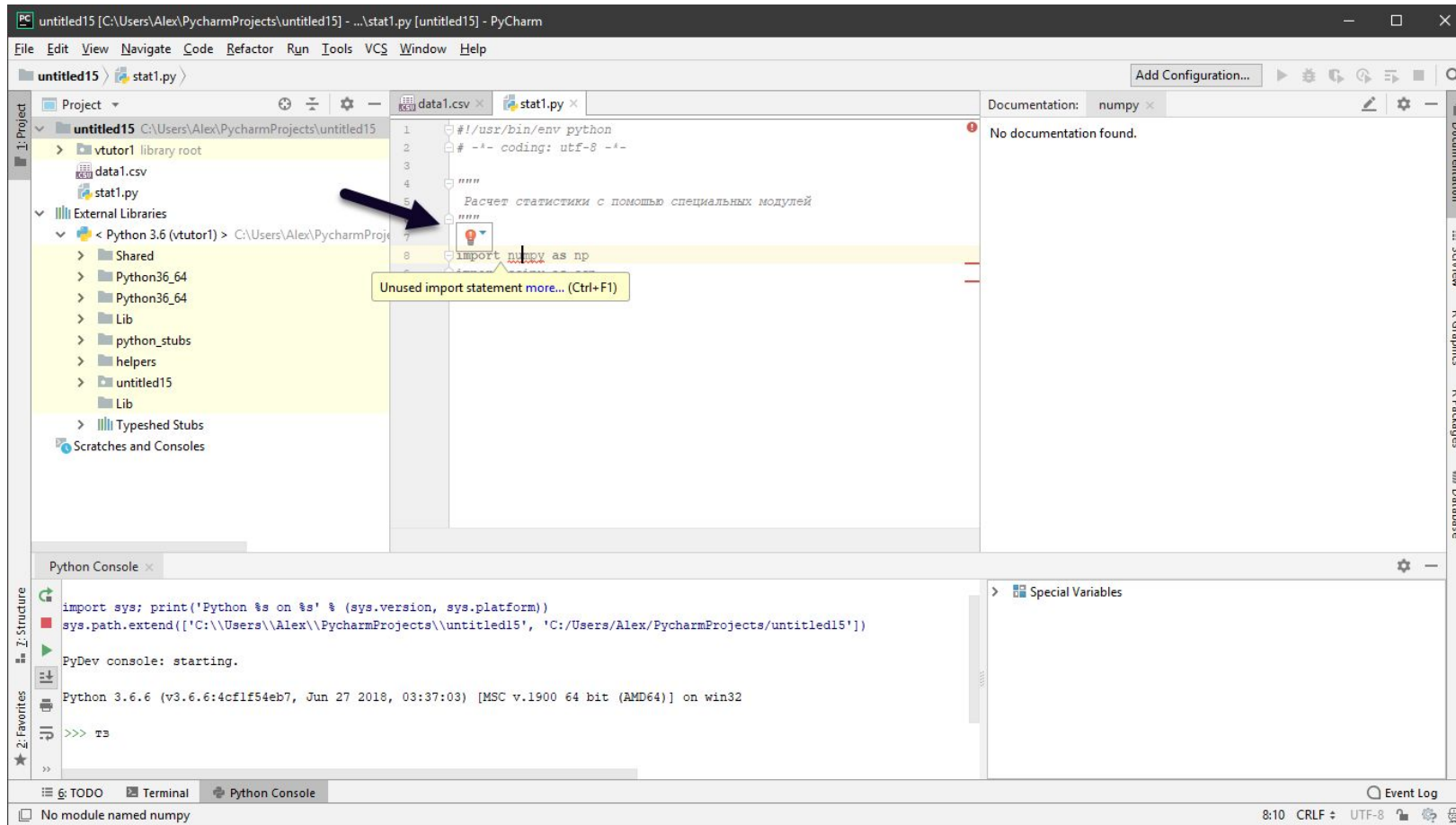
```
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\Users\Alex\PycharmProjects\untitled15', 'C:/Users/Alex/PycharmProjects/untitled15'])

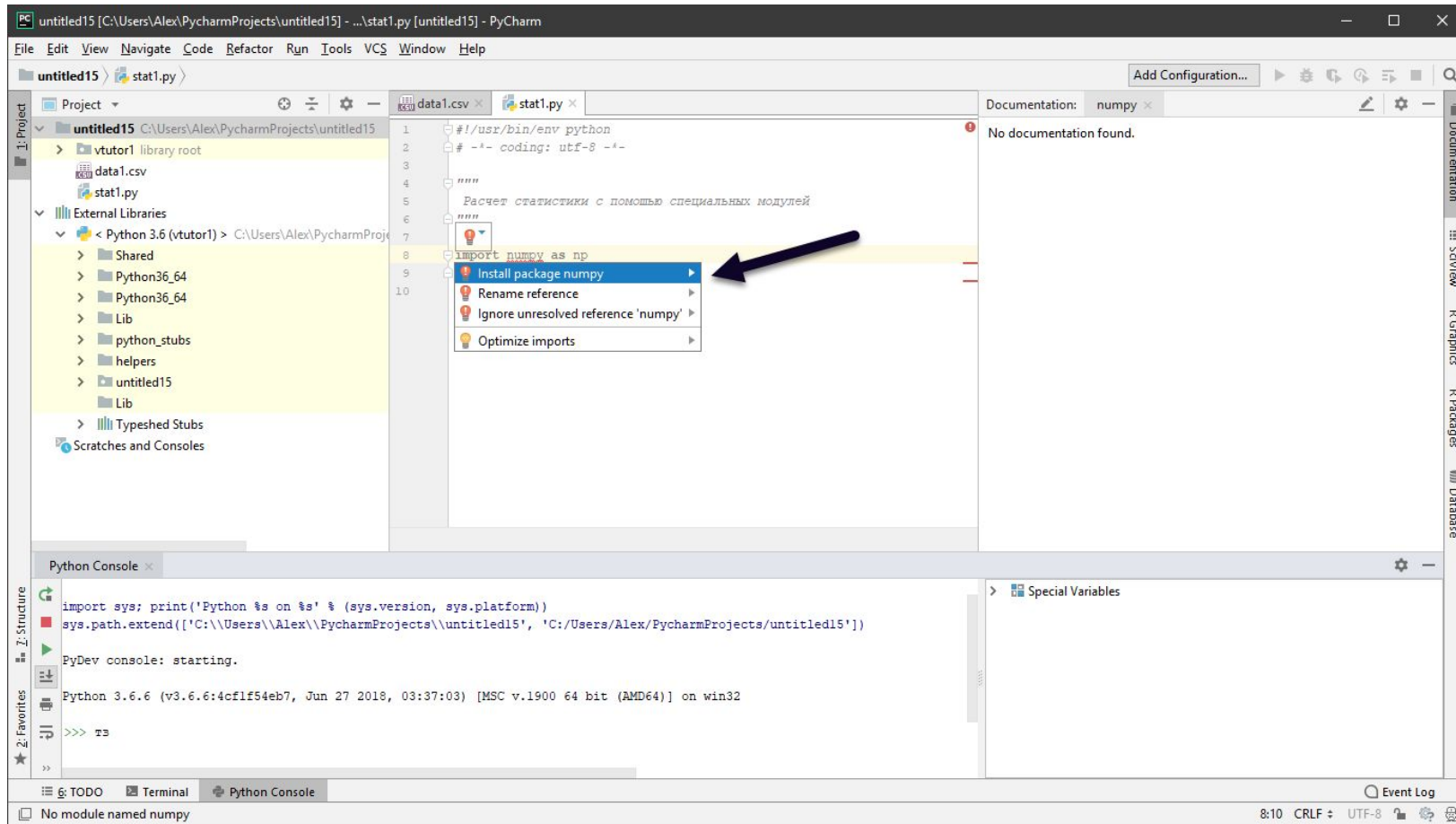
PyDev console: starting.

Python 3.6.6 (v3.6.6:4cf1f54eb7, Jun 27 2018, 03:37:03) [MSC v.1900 64 bit (AMD64)] on win32

>>> t3
>>>
```

The status bar at the bottom indicates the time is 10:1, the encoding is CRLF, and the file encoding is UTF-8.





The screenshot shows the PyCharm IDE interface. The main editor window displays a Python script named `stat1.py` with the following content:

```
1 #!/usr/bin/env python
2 #-*- coding: utf-8 -*-
3
4 """
5 Расчет статистики с помощью специальных модулей
6 """
7
8 import numpy as np
9 import scipy as scp
10
```

The `import scipy as scp` line is highlighted in yellow, and a red error icon is visible next to it. The right-hand pane shows the documentation for `scipy`, which displays "No documentation found." The bottom pane shows the Python Console with the following output:

```
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\Alex\\PycharmProjects\\untitled15', 'C:/Users/Alex/PycharmProjects/untitled15'])

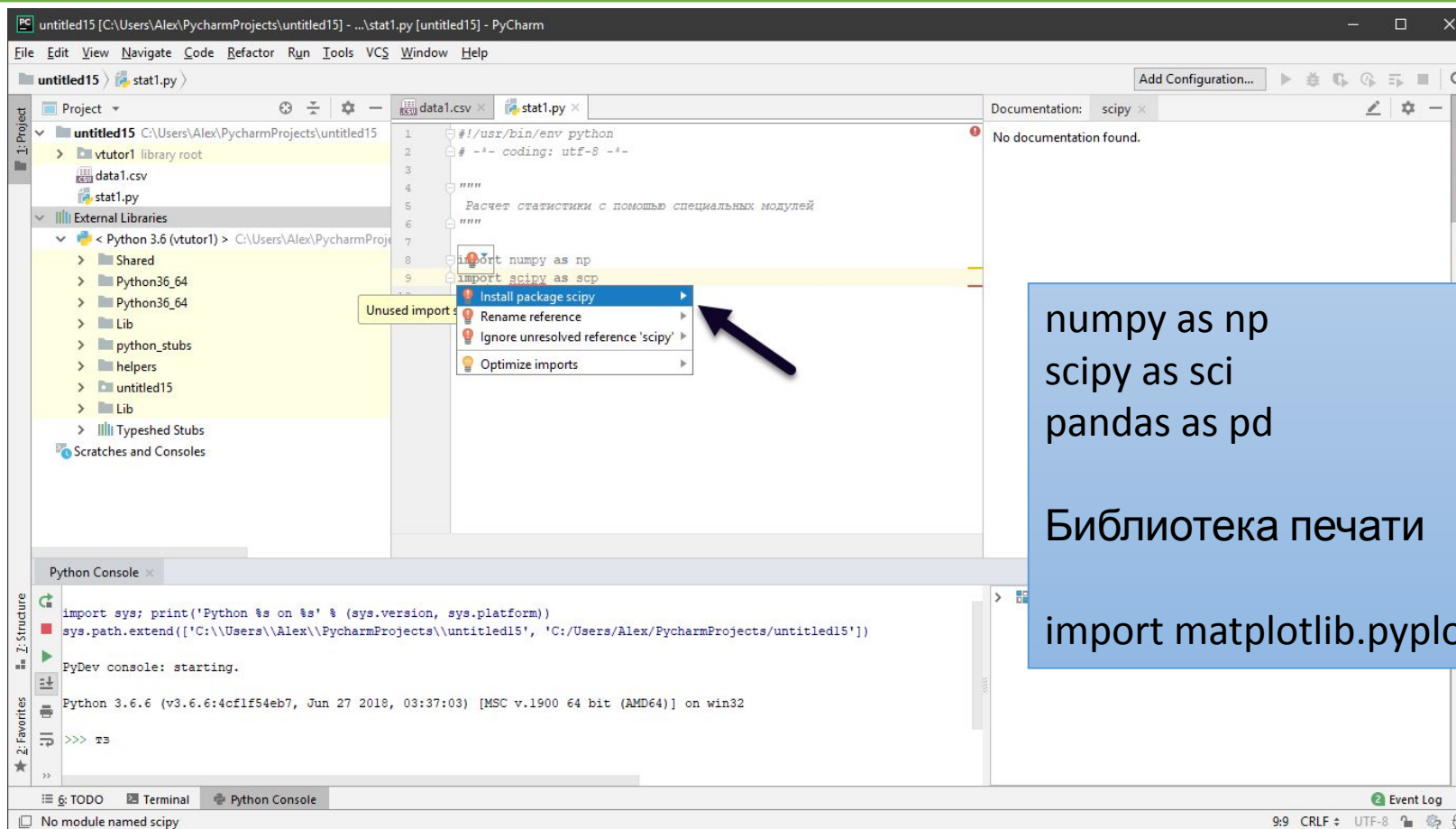
PyDev console: starting.

Python 3.6.6 (v3.6.6:4cf1f54eb7, Jun 27 2018, 03:37:03) [MSC v.1900 64 bit (AMD64)] on win32

>>> ts
>>>
```

The status bar at the bottom indicates the error: "No module named scipy". The system tray shows the time as 9:9, CRLF encoding, and UTF-8 font encoding.

Добавьте (загрузите модули при необходимости)



numpy as np
scipy as sci
pandas as pd

Библиотека печати

import matplotlib.pyplot as plt

Чтение данных

The screenshot displays the PyCharm IDE interface for a project named 'untitled15'. The main editor window shows the file 'stat1.py' with the following code:

```
1 #!/usr/bin/env python
2 # -*- coding: utf-8 -*-
3
4 """
5 Расчет статистики с помощью специальных модулей
6 """
7
8 # импорт научных библиотек
9 import numpy as np
10 import scipy as scp
11 import pandas as pd
12
13 # импорт библиотеки печати
14 from pylab import matplotlib
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 data = pd.read_csv("data1.csv")
19
20
21
22
23
```

The line `data = pd.read_csv("data1.csv")` is highlighted in yellow, with a red circle '1' and an arrow pointing to it. Below the code editor, the 'run' button is visible, with a red circle '2' next to it. To the right of the code editor, the 'Documentation' window shows the documentation for `pandas.core.reshape.pivot`. Below the documentation, the 'Special Variables' window is open, showing the variable `data = (DataFrame)` with a red circle '3' and an arrow pointing to it. The Python Console at the bottom shows the execution of the script:

```
Python Console x stat1 x
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\Users\Alex\PycharmProjects\untitled15', 'C:/Users/Alex/PycharmProjects/untitled15'])
PyDev console: starting.
Python 3.6.6 (v3.6.6:4c1f54eb7, Jun 27 2018, 03:37:03) [MSC v.1900 64 bit (AMD64)] on win32
>>> runfile('C:/Users/Alex/PycharmProjects/untitled15/stat1.py', wdir='C:/Users/Alex/PycharmProjects/untitled15')
>>>
>>>
```

The status bar at the bottom indicates that packages were installed successfully: 'matplotlib' (5 minutes ago). The system clock shows 18:32, and the encoding is UTF-8.

Что неверно?

- Разделитель – точка с запятой
- Десятичная точка – запятая
- Лучше так
- `data = pd.read_csv("data1.csv",sep=";",decimal=",")`

untitled15 [C:\Users\Alex\PycharmProjects\untitled15] - ...stat1.py [untitled15] - PyCharm

File Edit View Navigate Code Refactor Run Tools VCS Window Help

untitled15 stat1.py

Project: untitled15 C:\Users\Alex\PycharmProjects\untitled15

- vtutor1 library root
 - data1.csv
 - stat1.py
- External Libraries
 - Python 3.6 (vtutor1) > C:\Users\Alex\PycharmProj
 - Shared
 - Python36_64
 - python_stubs
 - helpers
 - untitled15
 - Lib
 - Typeshed Stubs
 - Scratches and Consoles

```
1 #!/usr/bin/env python
2 # -*- coding: utf-8 -*-
3
4 """
5 Расчет статистики с помощью специальных модулей
6 """
7
8 # импорт научных библиотек
9 import numpy as np
10 import scipy as scp
11 import pandas as pd
12
13 # импорт библиотеки печати
14 from pylab import matplotlib
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 data = pd.read_csv("data1.csv", sep=";", decimal=",")
19
20 print(data.x)
```

SciView: Data Plots

	N	x	y	z
0	1	-2.21000	-0.28924	-32.15000
1	2	-0.35000	-0.07608	-3.25000
2	3	-1.80000	0.05420	-24.00000
3	4	0.40000	0.00700	10.00000
4	5	-2.59000	-0.04095	-33.85000
5	6	1.36000	-0.16589	26.40000
6	7	2.35000	0.01689	42.25000
7	8	-2.89000	0.29906	-35.35000
8	9	0.35000	-0.00127	14.25000
9	10	0.59000	0.06090	18.85000
10	11	-2.40000	-0.02492	-25.00000
11	12	1.01000	0.11730	27.15000
12	13	-0.37000	-0.00332	7.45000
13	14	-2.74000	0.18902	-27.10000
14	15	1.69000	0.07509	40.35000
15	16	1.59000	-0.03788	39.85000
16	17	0.58000	-0.00855	25.70000

data Format: %

Special Variables

```
> data = (DataFrame) N x y z\n0 1-2.21...View as DataFrame
```

320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89

Name: x, Length: 326, dtype: float64

>>>

51 chars 18:1 CRLF UTF-8

Event Log

Packages installed successfully: Installed packages: 'matplotlib' (24 minutes ago)

untitled15 [C:\Users\Alex\PycharmProjects\untitled15] - ...stat1.py [untitled15] - PyCharm

File Edit View Navigate Code Refactor Run Tools VCS Window Help

untitled15 > stat1.py >

Project

- untitled15 C:\Users\Alex\PycharmProjects\untitled15
 - vtutor1 library root
 - data1.csv
 - stat1.py
 - External Libraries
 - < Python 3.6 (vtutor1) > C:\Users\Alex\PycharmProj
 - Shared
 - Python36_64
 - Python36_64
 - Lib
 - python_stubs
 - helpers
 - untitled15
 - Lib
 - Typeshed Stubs
 - Scratches and Consoles

data1.csv

```
5  Расчет статистики с помощью специальных модулей
6  """
7
8  # импорт научных библиотек
9  import numpy as np
10 import scipy as scp
11 import pandas as pd
12
13 # импорт библиотеки печати
14 from pylab import matplotlib
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 data = pd.read_csv("data1.csv", sep=";", decimal=",")
19
20 print(data.x)
21
22 plt.figure()
23 data.plot()
24 plt.show()
25
26
27
```

SciView: Data Plots

VG (24-bit color) 55.54 KB

Documentation
SciView
R Graphics
R Packages
Database

stat1(4) x stat1(5) x stat1(6) x stat1(7) x stat1(8) x stat1(9) x stat1(10) x stat1(11) x stat1(12) x stat1(13) x stat1(14) x stat1(15) x stat1(16) x stat1(17) x stat1(18) x

320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89

Name: x, Length: 326, dtype: float64

>>>

Special Variables

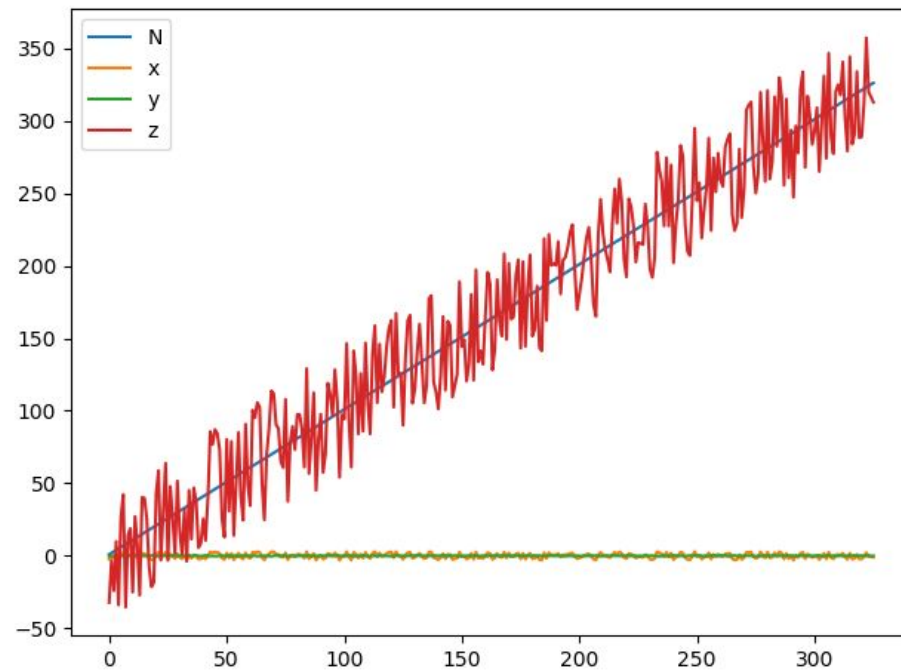
data = (DataFrame) N x y z\n0 1-221..View as DataFrame

Event Log

26:1 CRLF UTF-8

Packages installed successfully: Installed packages: 'matplotlib' (26 minutes ago)

<code>plt.figure()</code>	Создает место для графика (в режиме научных расчетов не обязательно)
<code>data.plot()</code>	Формирует график
<code>plt.show()</code>	Отображает график



The image shows a PyCharm IDE window with the following components:

- Project View:** Shows the project structure for 'untitled15', including 'data1.csv' and 'stat1.py'.
- Code Editor:** Contains the following Python code:

```
6 """
7
8 # импорт научных библиотек
9 import numpy as np
10 import scipy as scp
11 import pandas as pd
12
13 # импорт библиотеки печати
14 from pylab import matplotlib
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 fig = plt.figure()
19 data.plot()
20 plt.show()
21
22 plt.plot(data.x, data.y)
23 plt.show()
```
- SciView:** Displays a scatter plot of data points. The plot shows a dense cloud of blue points, with axes ranging from approximately -3 to 3 on the x-axis and -0.4 to 0.3 on the y-axis. The plot is titled 'stat1' and has a resolution of 'VG (24-bit color) 92.19 KB'.
- Console:** Shows the output of the script, including the data points and the variable 'x' with its length and data type.

```
320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89
Name: x, Length: 326, dtype: float64
>>>
```
- Bottom Panel:** Includes a 'Terminal' and 'Python Console' tab, and a status bar at the bottom indicating 'Packages installed successfully: Installed packages: 'matplotlib' (33 minutes ago)'. The system tray shows the time '27:10', encoding 'CRLF', and language 'UTF-8'.

- Такая структура, как фрейм данных pandas является близким аналогом фрейма из языка R
- В частности, он сам «знает» что и как надо напечатать
- Фрейм состоит из переменных (колонок), и строк
- Существуют механизмы выбора отдельных колонок или их множества
- Существуют инструменты отбора данных в колонках (например по условию)
- Потенциально существует возможность поменять местами строки и столбцы (транспонировать таблицу)

The image shows a PyCharm IDE window titled "untitled15 [C:\Users\Alex\PycharmProjects\untitled15] - ...\stat1.py [untitled15] - PyCharm". The interface includes a menu bar (File, Edit, View, Navigate, Code, Refactor, Run, Tools, VCS, Window, Help), a toolbar, and several panels:

- Project View:** Shows the project structure for "untitled15", including "vtutor1" (library root, data1.csv, stat1.py) and "External Libraries" for Python 3.6.
- Code Editor:** Contains the following Python code:

```
11 import pandas as pd
12
13 # импорт библиотеки печати
14 from pylab import matplotlib
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 data = pd.read_csv("data1.csv", sep=";", decimal=",")
19
20 print(data.x)
21
22 plt.plot(data.x)
23 plt.show()
24
25 data.hist()
26 plt.show()
```
- SciView:** Displays four histograms for variables N, x, y, and z. The histograms show the distribution of data points for each variable.
- Console:** Shows the output of the script, including the values of data.x and the type of the variable x.

```
320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89
Name: x, Length: 326, dtype: float64
>>>
```
- Special Variables:** Shows the variable "data" as a DataFrame with columns N, x, y, z.

A black circle highlights the code lines `data.hist()` and `plt.show()` in the code editor. A black arrow points from this circle to the histograms in the SciView panel, indicating the relationship between the code and the resulting plots.

At the bottom of the IDE, there is a status bar with the text "Packages installed successfully: Installed packages: 'matplotlib' (34 minutes ago)" and system information "32:1 CRLF UTF-8".

The image shows a PyCharm IDE window with the following components:

- Project View:** Shows a project named 'untitled15' with files 'data1.csv' and 'stat1.py'.
- Code Editor:** Contains the following Python code:

```
12  
13 # импорт библиотеки печати  
14 from pylab import matplotlib  
15 # импорт библиотеки графики  
16 import matplotlib.pyplot as plt  
17  
18 data = pd.read_csv("data1.csv", sep=";", decimal=",")  
19  
20 print(data.x)  
21  
22 plt.figure()  
23 data.plot()  
24  
25 data.hist()  
26 plt.show()
```
- Console:** Shows the output of the script:

```
320 -2.15  
321 -0.68  
322 2.27  
323 -0.29  
324 -0.58  
325 -0.89  
Name: x, Length: 326, dtype: float64  
>>>
```
- SciView:** Displays a histogram plot of the data. The x-axis ranges from -3 to 3, and the y-axis ranges from 0 to 40. The plot shows a distribution of data points with a peak around -2. The plot title is 'NG (24-bit color) 8.55 KB'.
- Bottom Panel:** Shows the 'Special Variables' section with the variable 'data' defined as a DataFrame with columns 'x', 'y', and 'z'. The DataFrame has 1 row and 3 columns.

A blue circle highlights the code lines `data.hist()` and `plt.show()` in the code editor. A black arrow points from this circle to the histogram plot in the SciView panel.

The image shows a PyCharm IDE window with a Python script and a histogram plot. The script in the center pane reads:

```
15 # импорт библиотеки графики
16 import matplotlib.pyplot as plt
17
18 data = pd.read_csv("data1.csv", sep=";", decimal=",")
19
20 print(data.x)
21
22 plt.figure()
23 data.plot()
24 plt.show()
25
26 data.y
```

A callout box highlights the following code:

```
data.x.hist()
plt.show()

data["y"].hist()
plt.show()
```

The right pane displays a histogram plot of the data. The x-axis ranges from -0.4 to 0.3, and the y-axis ranges from 0 to 70. The plot shows a distribution of data points with a peak around 0.0. The SciView panel on the right shows a list of plots, including the current histogram.

The bottom pane shows the Python Console output:

```
320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89
Name: x, Length: 326, dtype: float64
>>>
```

The bottom status bar indicates that the packages installed successfully: 'matplotlib' (37 minutes ago). The system tray shows the time as 36:10, the encoding as CRLF, and the font size as UTF-8.

The image shows a PyCharm IDE window with a Python script named `stat1.py` and its execution results. The script contains the following code:

```
plt.show()
plt.plot(data["x"], data["y"])
plt.hist()
plt.show()
data["z"].hist()
plt.show()
plt.scatter(data["N"], data["z"])
plt.show()
data.loc[:, ["x", "y", "z"]].hist()
plt.show()
```

The output of the script is displayed in the SciView window, showing three histograms: one for variable `x`, one for variable `y`, and one for variable `z`. The `x` histogram has a range from approximately -2 to 2, the `y` histogram from -0.4 to 0.2, and the `z` histogram from 0 to 300. The `z` histogram shows a distribution with a peak around 100. The SciView window also displays a table of special variables:

Special Variables	N	x	y	z
data = (DataFrame)	1-221..	View as DataFrame		

The console window shows the output of the `print` statement, displaying the values of `x` for each row:

```
320 -2.15
321 -0.68
322 2.27
323 -0.29
324 -0.58
325 -0.89
Name: x, Length: 326, dtype: float64
```

The status bar at the bottom indicates that the packages `matplotlib` and `matplotlib` were installed successfully 50 minutes ago.

The image shows a PyCharm IDE window titled "untitled15 [C:\Users\Alex\PycharmProjects\untitled15] - ...\stat1.py [untitled15] - PyCharm". The main editor displays a Python script with the following code:

```
data.hist()  
plt.show()  
  
plt.scatter(data["N"], data["z"])  
plt.show()  
  
data.loc[:, ["x", "y", "z"]].hist()  
plt.show()  
  
data.loc[:, ["x", "z"]].plot.box()  
plt.show()
```

A blue circle highlights the last two lines of code. A blue box highlights the line `data.loc[:, ["x", "z"]].plot.box()`, with an arrow pointing to a SciView plot window. The SciView window shows a box plot for variables 'x' and 'z'. The y-axis ranges from 0 to 350. Variable 'x' has a very narrow distribution near 0, while variable 'z' has a wider distribution with a median around 150. The SciView window also shows a histogram for 'z' and a scatter plot for 'N' vs 'z'.

The bottom console shows the following output:

```
stat1(29) x  stat1(30) x  stat1(31) x  
320 -2.15  
321 -0.68  
322 2.27  
323 -0.29  
324 -0.58  
325 -0.89  
Name: x, Length: 326, dtype: float64  
>>>
```

The bottom status bar indicates "Packages installed successfully: Installed packages: 'matplotlib' (today 1:17)".

```
data["z"].hist()
plt.show()

plt.scatter(data["N"], data["z"])
plt.show()

data.loc[:, ["x", "y", "z"]].hist()
plt.show()

data.loc[:, ["x", "z"]].plot.box()
plt.show()

print("Stat")
print("average".center(64, "="))
print(data.median())
```

SciView: Data Plots

'NG (24-bit color) 7.01 KB

Documentation
SciView
R Graphics
R Packages
Database

stat1(40) × stat1(41) × stat1(42) × stat1(43) × stat1(44) × 31

Special Variables
data = (DataFrame) N x y z\n0 1-41...View as DataFrame

Structure

Stat	
N	163.500000
x	-0.085000
y	0.018796
z	162.250000
dtype:	float64

Python Console

Low Memory: The IDE is running low on memory and this might affect performance. Please consider increasing available heap. // Configure (2 minutes ago)

45:11 CRLF UTF-8

Расчет отдельных показателей

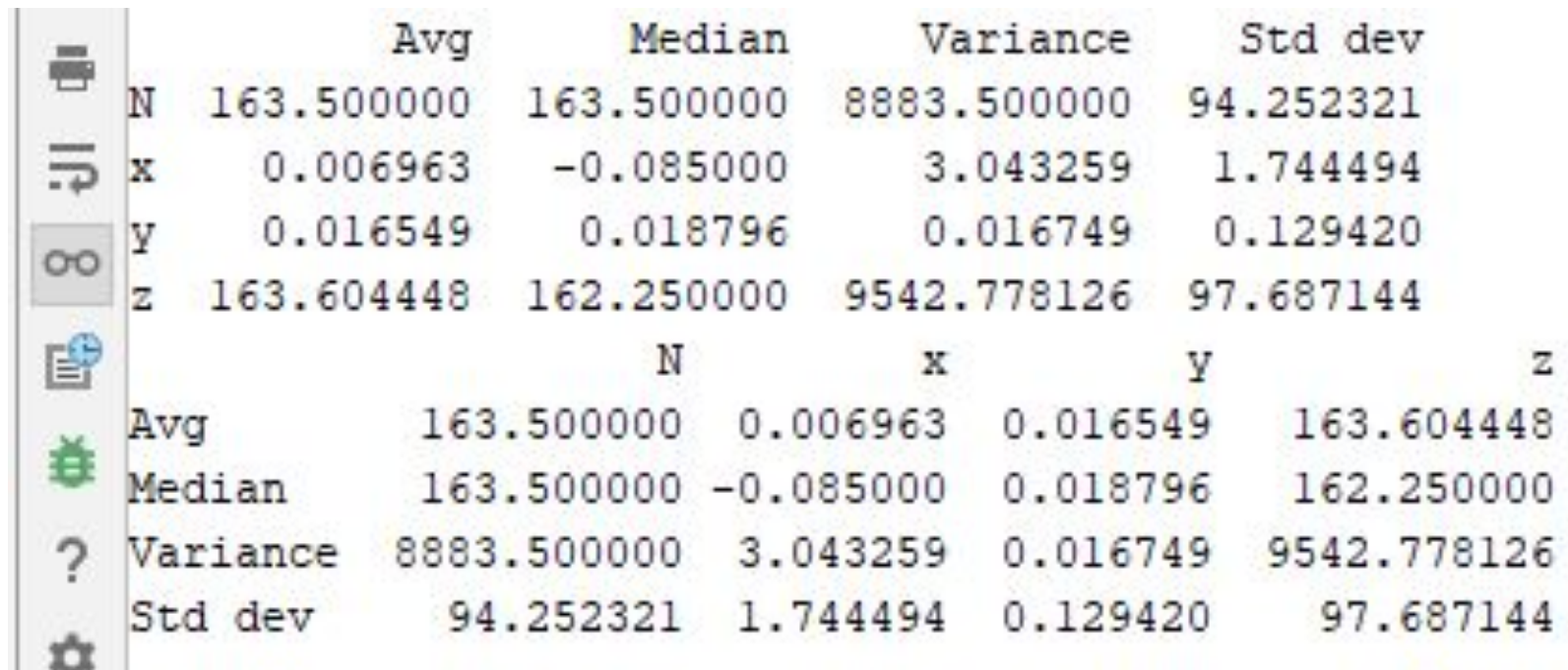
- `print("Stat")`
- `print("average".center(64,"="))`
- `print(data.median())`
- `print("variance".center(64,"="))`
- `print(data.var())`
- `print("std dev".center(64,"="))`
- `print(data.std())`

Специальная таблица под статистику

- `stat = pd.DataFrame()`
- `stat["Avg"] = data.mean()`
- `stat["Median"] = data.median()`
- `stat["Variance"] = data.var()`
- `stat["Std dev"] = data.std()`
- `print(stat)`
- `print(stat.T)`

Вывод на экран как в «прямом», так и «развернутом виде»

- `print(stat)`
- `print(stat.T)` – транспонированная таблица



		Avg	Median	Variance	Std dev
N	163.500000	163.500000	8883.500000	94.252321	
x	0.006963	-0.085000	3.043259	1.744494	
y	0.016549	0.018796	0.016749	0.129420	
z	163.604448	162.250000	9542.778126	97.687144	
		N	x	y	z
Avg	163.500000	0.006963	0.016549	163.604448	
Median	163.500000	-0.085000	0.018796	162.250000	
Variance	8883.500000	3.043259	0.016749	9542.778126	
Std dev	94.252321	1.744494	0.129420	97.687144	