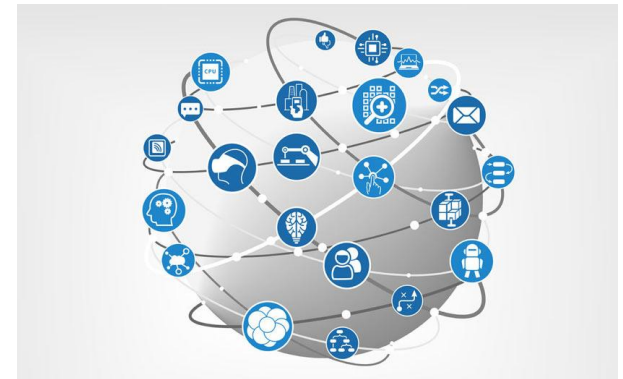
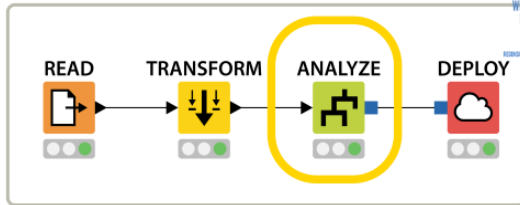
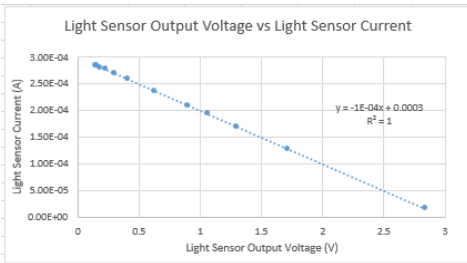


# Predictive Analytics and Machine Learning Basics

## Class 2: Creating a Predictive Analytic Model



July 17, 2018  
Don Wilcher

# Class 2: Creating a Predictive Analytic Model



## Agenda

- The Predictive Analytics Process
- What is Linear Regression
- Exploring Orange
- Lab Project: Predicting the electrical behavior of a light sensor circuit.

# The Predictive Analytics Process



There are seven steps to the Predictive Analytics Process.

- 1. Define Project**
- 2. Data Collection**
- 3. Data Analysis**
- 4. Statistics**
- 5. Modeling**
- 6. Deployment**
- 7. Model Monitoring**

# The Predictive Analytics Process...



- 1. Project Definition** – Identifying the project outcomes based on the design objectives specified.
- 2. Data Collection** – Acquiring and preparing data for the purpose of predicting future trends for processing and monitoring systems.
- 3. Data Analysis** – The process of reviewing data results using statistical methods.

# The Predictive Analytics Process...



**4. Statistics** – The validation of assumptions and hypothesis using statistical methods.

**5. Modeling** – The ability to create predictive models that can accurately forecast future trends.

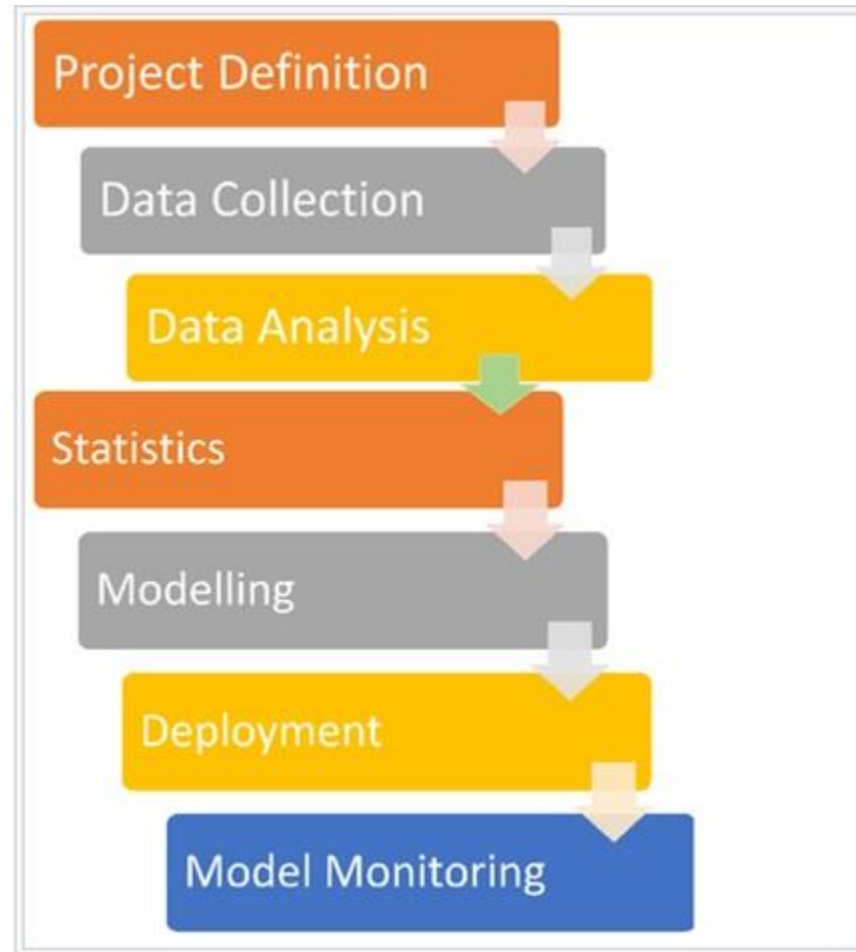
**6. Deployment** – To send out the predictive model for use in decision making processes, tasks and activities.

**7. Model Monitoring** – The management and monitoring the predictive performance of the analytical model.

# The Predictive Analytics Process...



## Predictive Analytics (PA) Process Model



# What is Linear Regression?



**Linear Regression** – Identifying the project outcomes based on the design objectives specified.

## FAQs:

- The focus is on predicting a value of  $Y$  having been given the value of  $X$ .
- The regression line is called the *prediction line*.
- The regression line is a straight line that lies closet to all the points in the scatter plot.

### Source:

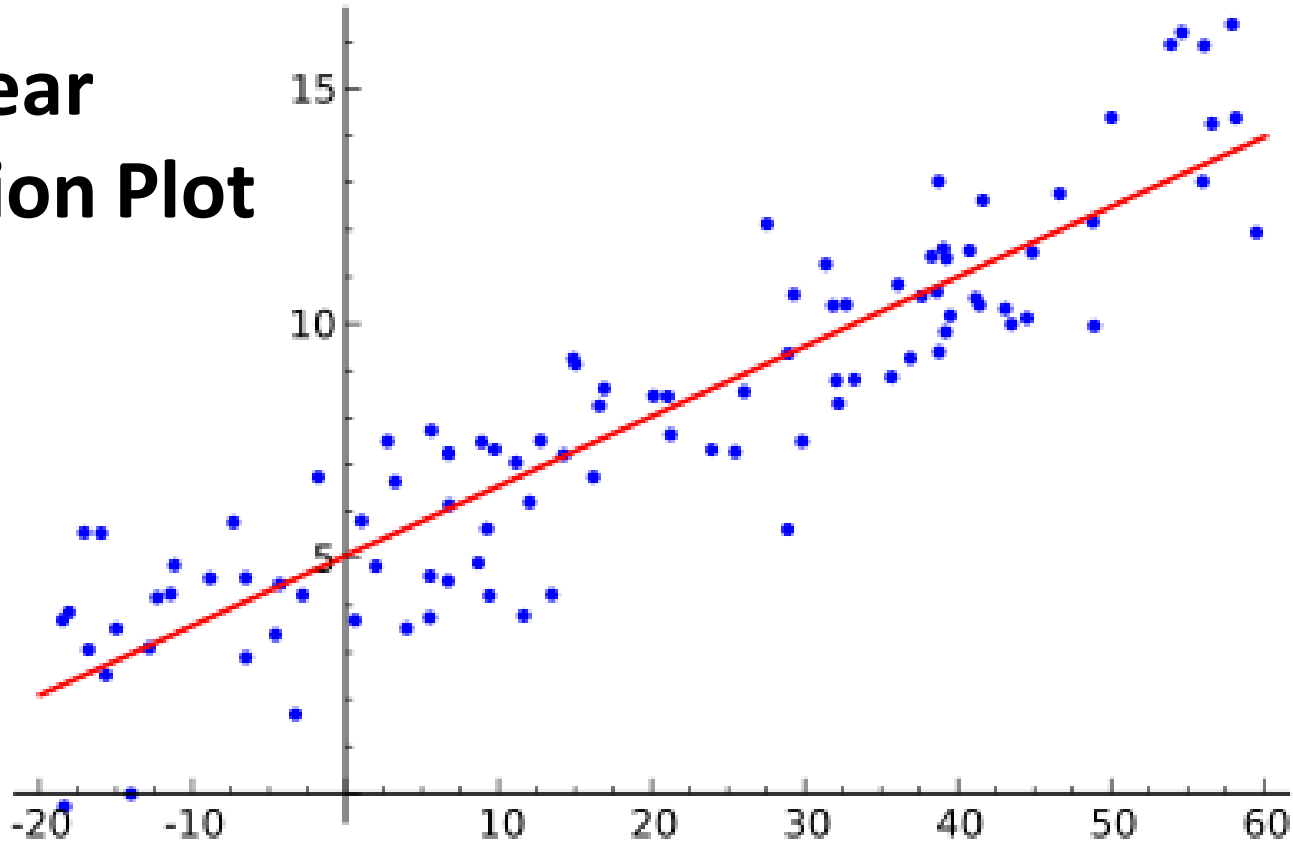
Sprinthall, R. (1987). Basic statistical analysis, 3<sup>rd</sup> ed. Englewood Cliffs, NJ: Prentice Hall Inc.

Presented by:

# What is Linear Regression? . . .



## Linear Regression Plot





# Question 1



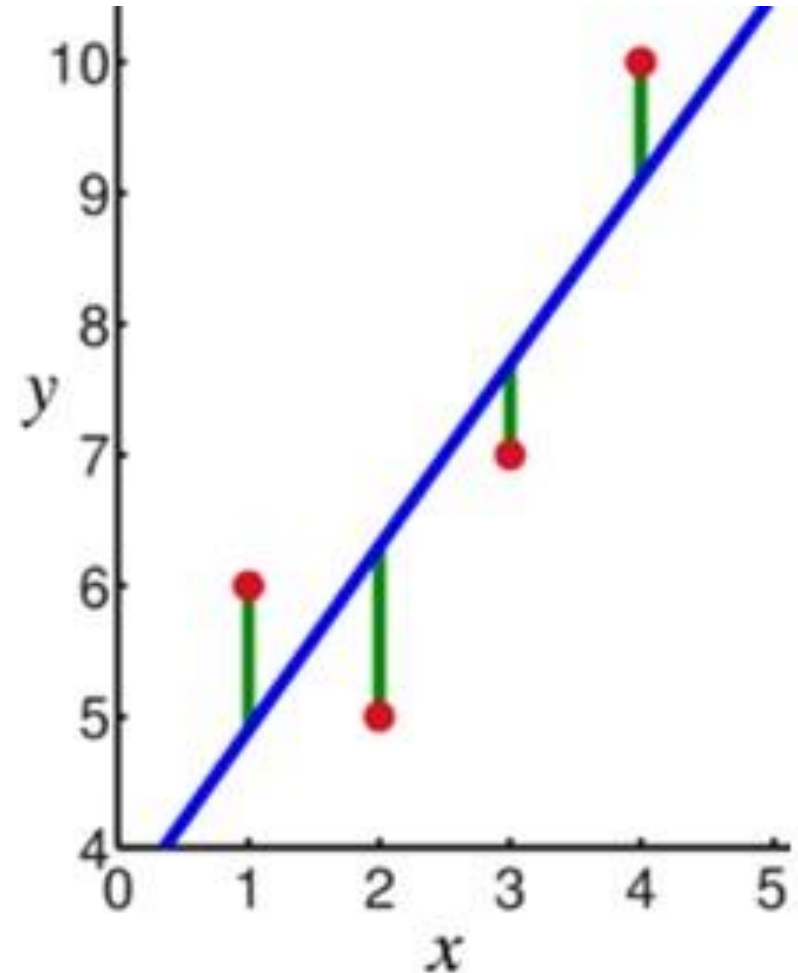
**True or False: There are eight steps for the Predictive Analytics Model process.**

# What is Linear Regression? . . .



## Linear Regression Plot Anatomy

- a) Red dots: observation
- b) Green line: deviation
- c) Blue line:  $y$  vs  $x$



# What is Linear Regression? . . .



## The Extent of the Scatter Around the Regression

- The closer the points on the scatter plot cluster around the regression line, the higher is the resulting correlation between  $x$  and  $y$ .
- The closer the points to the regression line, the more accurate is the resulting prediction.
- The higher the correlation, the closer the scatter points cluster around the regression line.

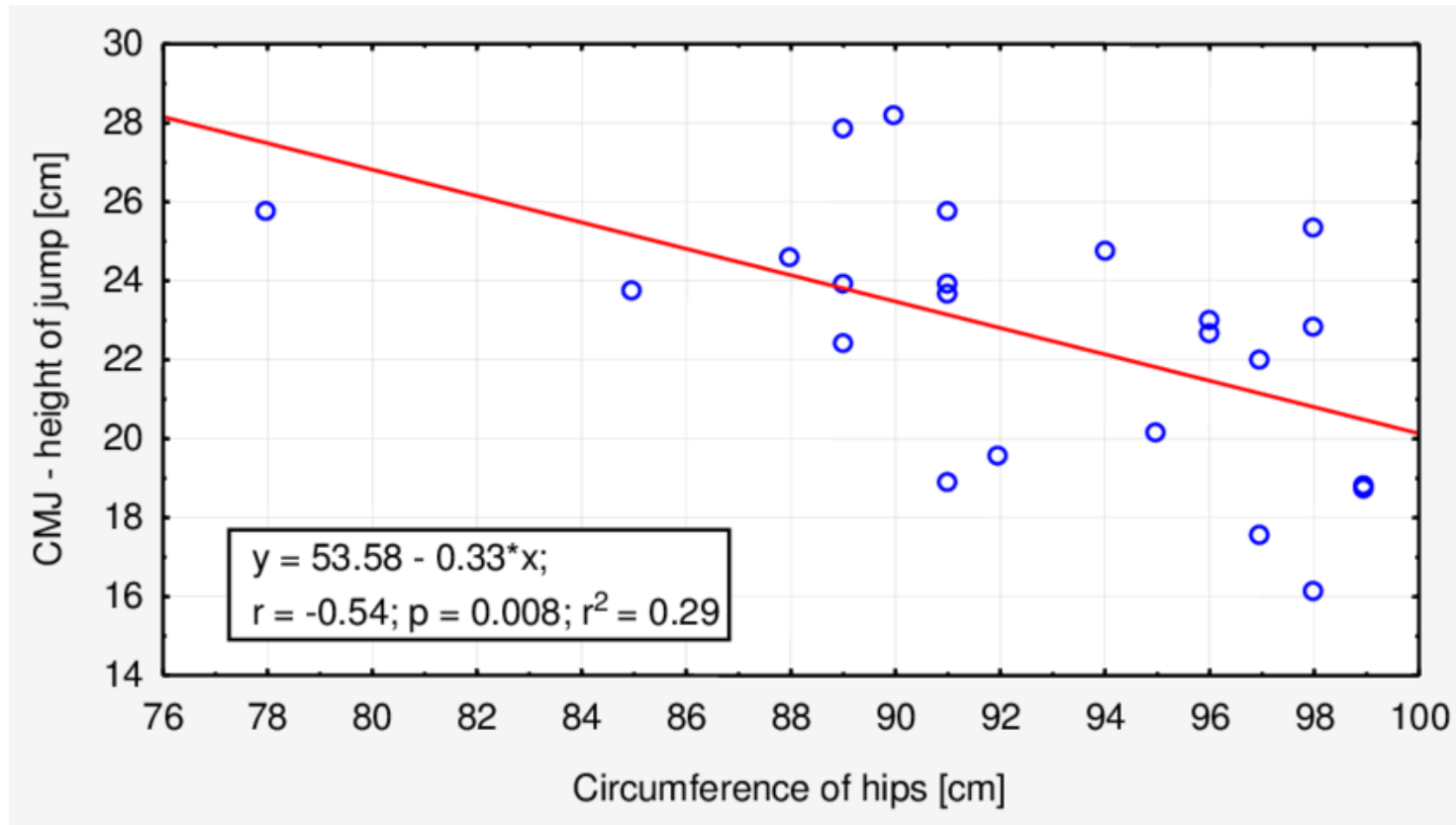
**Source:**

Sprinthall, R. (1987). Basic statistical analysis, 3<sup>rd</sup> ed. Englewood Cliffs, NJ: Prentice Hall Inc.

Presented by:



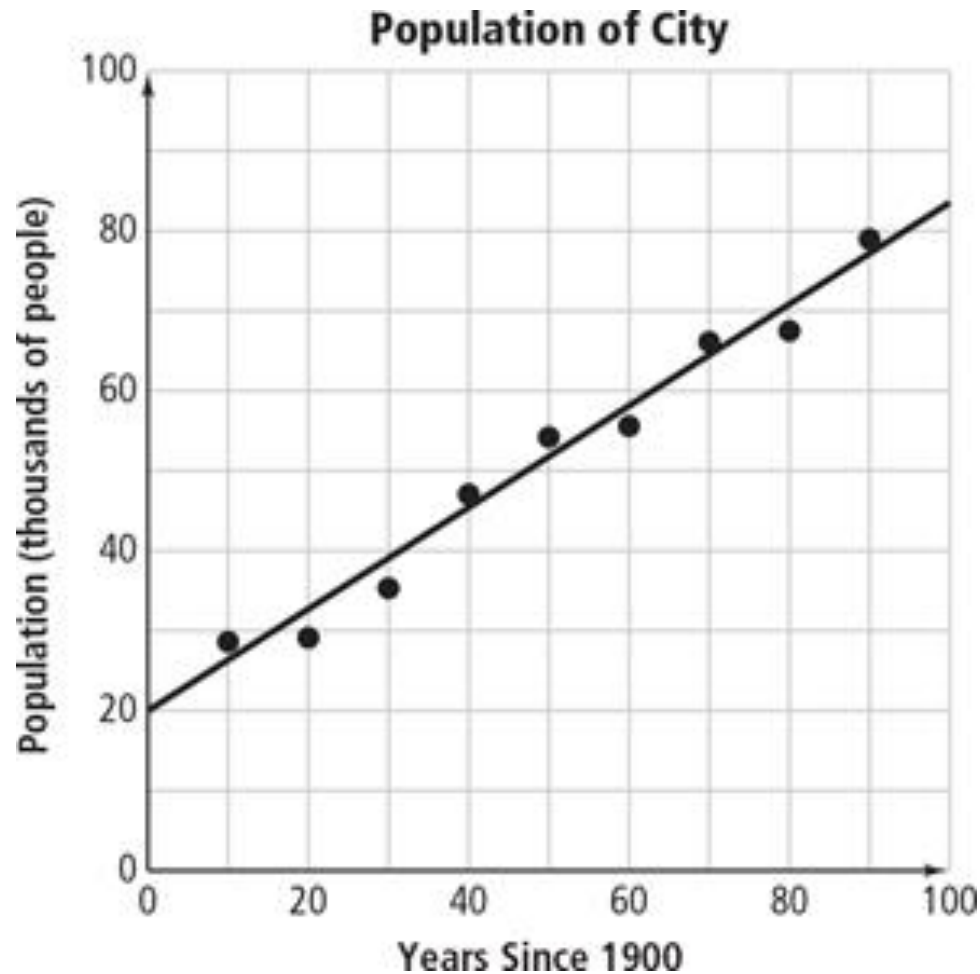
# What is Linear Regression? . . .



Source:

[https://www.researchgate.net/publication/309526456\\_Application\\_of\\_the\\_index\\_of\\_balance-stiffness\\_for\\_evaluation\\_of\\_the\\_process\\_of\\_maintaining\\_body\\_balance/figures?lo=1](https://www.researchgate.net/publication/309526456_Application_of_the_index_of_balance-stiffness_for_evaluation_of_the_process_of_maintaining_body_balance/figures?lo=1)

# What is Linear Regression? . . .



[http://flipbooks.pearsonschool.com/texasreview/mathematics/digits/TX\\_Digits\\_HomeworkHelper\\_HTML\\_Files/Grade%208/Volume%202/page\\_385.html](http://flipbooks.pearsonschool.com/texasreview/mathematics/digits/TX_Digits_HomeworkHelper_HTML_Files/Grade%208/Volume%202/page_385.html)

## Question 2



**True or False: Linear Regression is a complex Machine Learning Model for Predictive Analytics.**

# What is Linear Regression? . . .



Linear Regression based on this simple equation.

$$y = mx + b$$

output — slope — input — y-intercept

## Note:

and are coefficients for the linear equation

Source:

[https://www.sas.com/en\\_us/insights/analytics/predictive-analytics.html](https://www.sas.com/en_us/insights/analytics/predictive-analytics.html)

# Exploring Orange



Features

Screenshots

Download

Docs

Blog

Training

Donate

## Data Mining Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

Download Orange



Source:

<https://orange.biolab.si/>



# Exploring Orange. . .



orange

Features

Screenshots

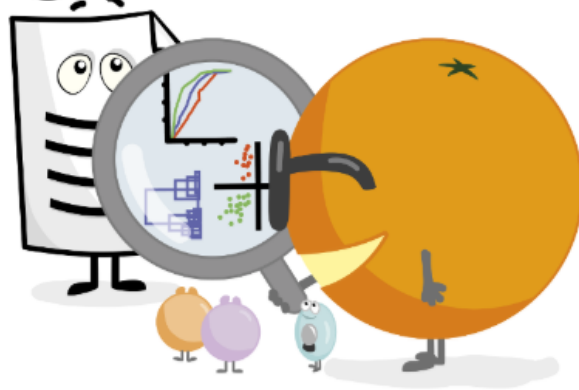
Download

Docs

Blog

Training

Donate



## Interactive Data Visualization

Perform simple data analysis with clever data visualization. Explore statistical distributions, box plots and scatter plots, or dive deeper with decision trees, hierarchical clustering, heatmaps, MDS and linear projections. Even your multidimensional data can become sensible in 2D, especially with clever attribute ranking and selections.

[Learn More](#)

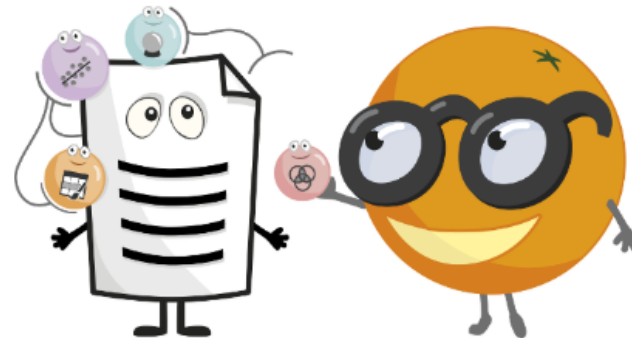
## Visual Programming

Interactive data exploration for rapid qualitative analysis with clean visualizations. Graphic user interface allows you to focus on exploratory data analysis instead of coding, while clever defaults make fast prototyping of a data analysis workflow extremely easy. Place widgets on the canvas, connect them, load your datasets and harvest the insight!

[Learn More](#) [Watch Video](#)

Source:

<https://orange.biolab.si/>



# Exploring Orange. . .



Example Workflows

## Example Workflows

### File and Data Table

The basic data mining units in Orange are called widgets. There are widgets for reading the data, preprocessing, visualization, clustering, classification and others. Widgets communicate through channels. Data mining workflow is thus a collection of widgets and communication channels.

In this workflow, there is a File widget that reads the data. File widget communicates this data to Data Table widget that shows the data spreadsheet. Notice how the output of the file widget is connected to the input of the Data Table widget. In Orange, the outputs of the widgets are on the right, and the inputs on the left of the widget.

A File widget. Double click to open it and select the data set file.

A Data Table widget. Double click the icon to see the data in a spreadsheet.

The output of the Data Table to send out any data (rows) that are selected to the widget.

This output is not used, hence dashed line. You can add another Data Table by clicking on its icon from the toolbox on the left, connect the output of Data Table to the input of new Data Table (1) and check if the selected data from Data Table is indeed sent to the downstream widget. This demo works best if both widgets are open, that is, their windows displayed.

The output of the File widget.

The input of the Data Table widget.

The communication channel. It passes the data set from the File widget to the Data Table.

Path: C:\Anaconda3\lib\site-packages\Orange\canvas\application\workflows\110-file-and-data-table-widget.ows

File and Data Table    Interactive Visualizations    Visualization of Data Subsets    Classification Tree    Principal Component Analysis    Hierarchic

# Exploring Orange. . .



Home   Screenshot   Download   Docs   Blog   Training   [Donate](#)

## Documentation



### Visual Programming

- Getting started
- YouTube tutorials
- Loading your data
- Widget catalog



### Development

- Widget development
- Example add-on



### Python Library

- Tutorial
- Reference
- Orange 2.7 documentation

Source:

<https://orange.biolab.si/docs/>

## Question 3



**Orange is an open source machine learning and \_\_\_\_\_ visualization tool for novice and expert developers.**

# Exploring Orange. . .



Home Screenshots Download Docs Blog Training [Donate](#)

Quick Search

## Data

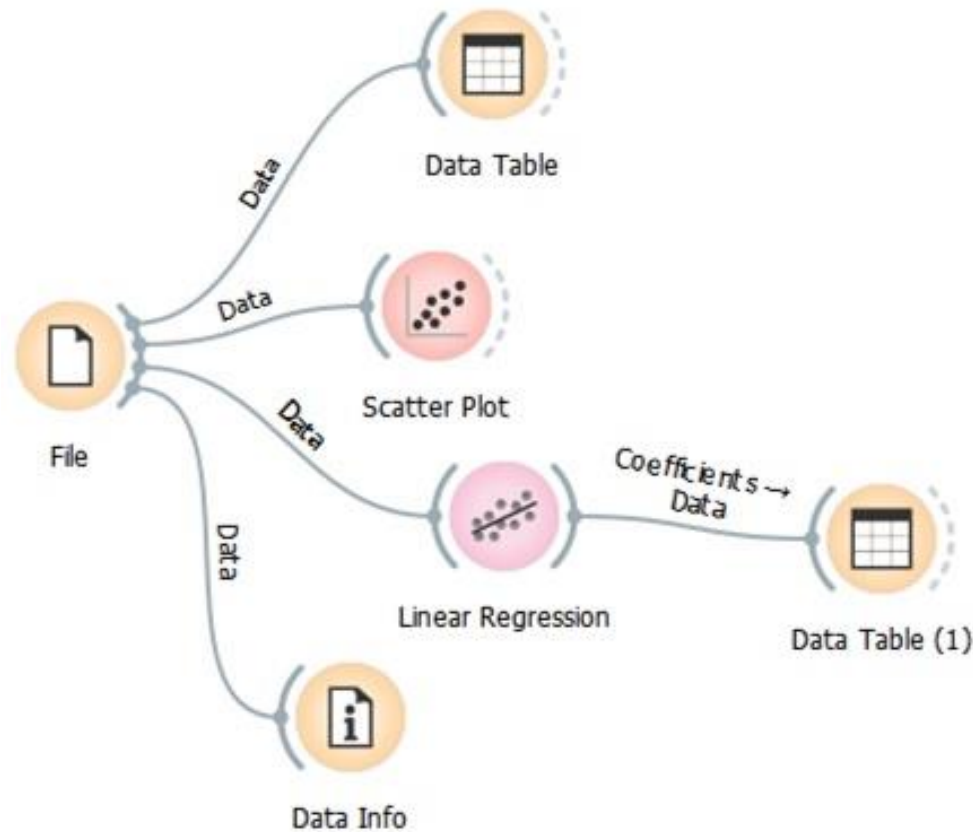
File	Datasets	SQL Table	Data Table	Paint Data	Data Info	Data Sampler
Select Columns	Select Rows	Rank	Merge Data	Concatenate	Transpose	Randomize
Preprocess	Impute	Outliers	Edit Domain	Python Script	Color	Continuize
Create Class	Discretize	Feature Constructor	Purge Domain	Save Data		

Source:  
<https://orange.biolab.si/toolbox/>

# Exploring Orange. . .



## Analyzing a Cricket's Chirp Rate vs Temperature: Workflow



# Exploring Orange. . .



## Cricket's Chirp Rate vs Temperature Data:

### CSV file

```
Chirp Rate, Temperature
20, 88.6
16, 71.6
19.8, 93.3
18.4, 84.3
17.1, 80.6
15.5, 75.2
14.7, 69.7
17.1, 82
15.4, 69.4
16.3, 83.3
15, 79.6
17.2, 82.6
16, 80.6
17, 83.5
14.4, 76.3
```

## Question 4



**Using slide 23: In configuring the input file, what feature is used as the target for the machine learning model in Orange?**



# Exploring Orange. . .



Data Table

Info  
15 instances (no missing values)  
1 feature (no missing values)  
Continuous target variable (no missing values)  
No meta attributes

Variables  
 Show variable labels (if present)  
 Visualize continuous values  
 Color by instance classes

Selection  
 Select full rows

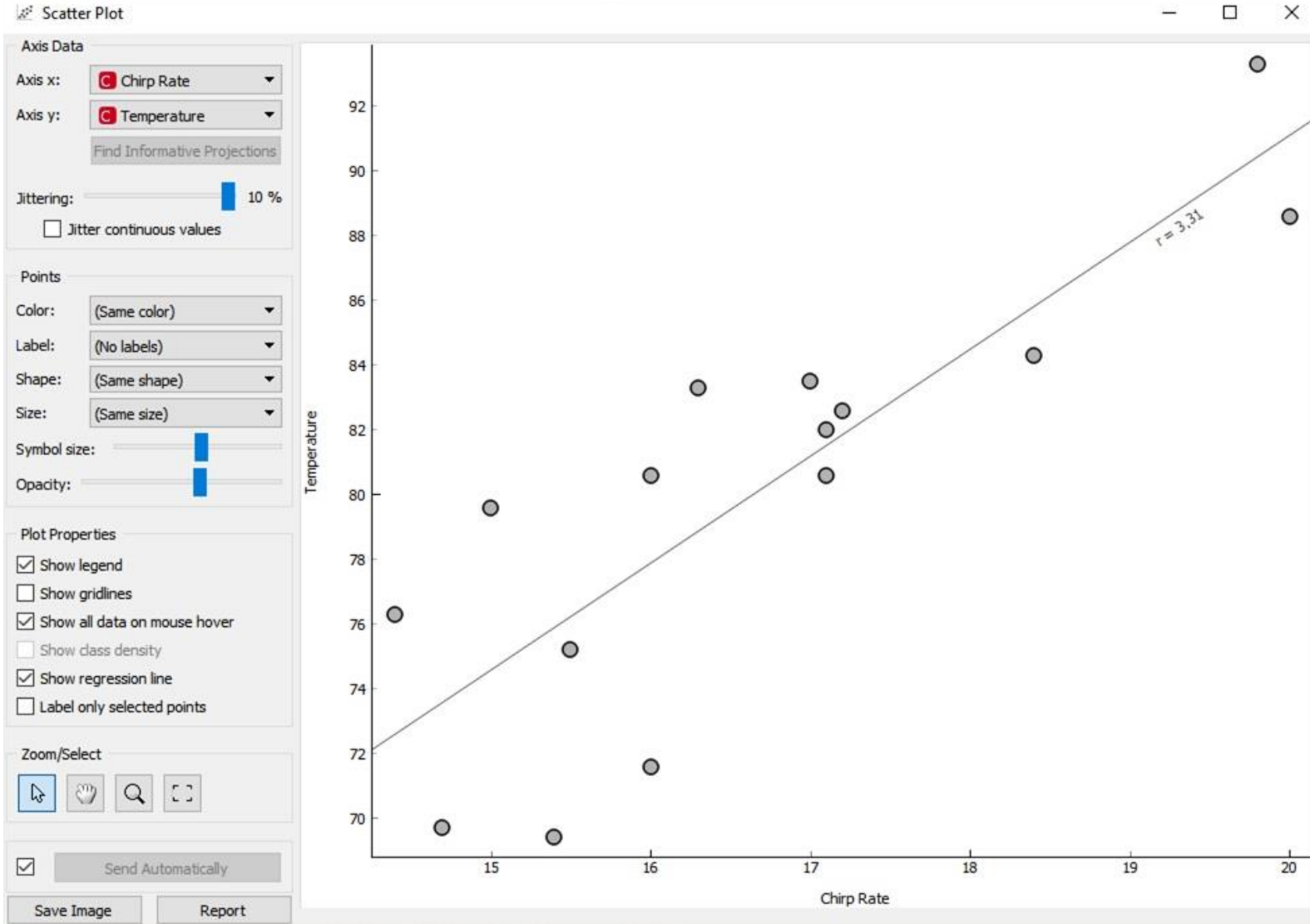
	Temperature	Chirp Rate
1	88.600	20.000
2	71.600	16.000
3	93.300	19.800
4	84.300	18.400
5	80.600	17.100
6	75.200	15.500
7	69.700	14.700
8	82.000	17.100
9	69.400	15.400
10	83.300	16.300
11	79.600	15.000
12	82.600	17.200
13	80.600	16.000
14	83.500	17.000
15	76.300	14.400

Restore Original Order

Report

Send Automatically

# Exploring Orange. . .



# Exploring Orange. . .



Data Table (1)

Info

- 2 instances (no missing values)
- 1 feature (no missing values)
- No target variable.
- 1 meta attribute (no missing values)

Variables

- Show variable labels (if present)
- Visualize continuous values
- Color by instance classes

Selection

- Select full rows

Restore Original Order

Report

Send Automatically

	name	coef
1	intercept	24.9660144
2	Chirp Rate	3.3057614

**Linear Regression equation.**

$$y = 3.3057614x + 24.9660144$$

# Exploring Orange. . .



**Predicting a new temperature with a future chirp rate example.**

**Given:**

$$x = 21Hz$$

$$y = 3.3057614x + 24.9660144$$

$$y = 3.3057614(21) + 24.9660144$$

$$y = 94.3869744 \text{ } ^\circ F$$



## Question 5

**Using slide 28: The numbers in the linear regression expression are the \_\_\_\_\_ of the equation.**

# Predicting the Electrical Behavior of a Light Sensor Circuit



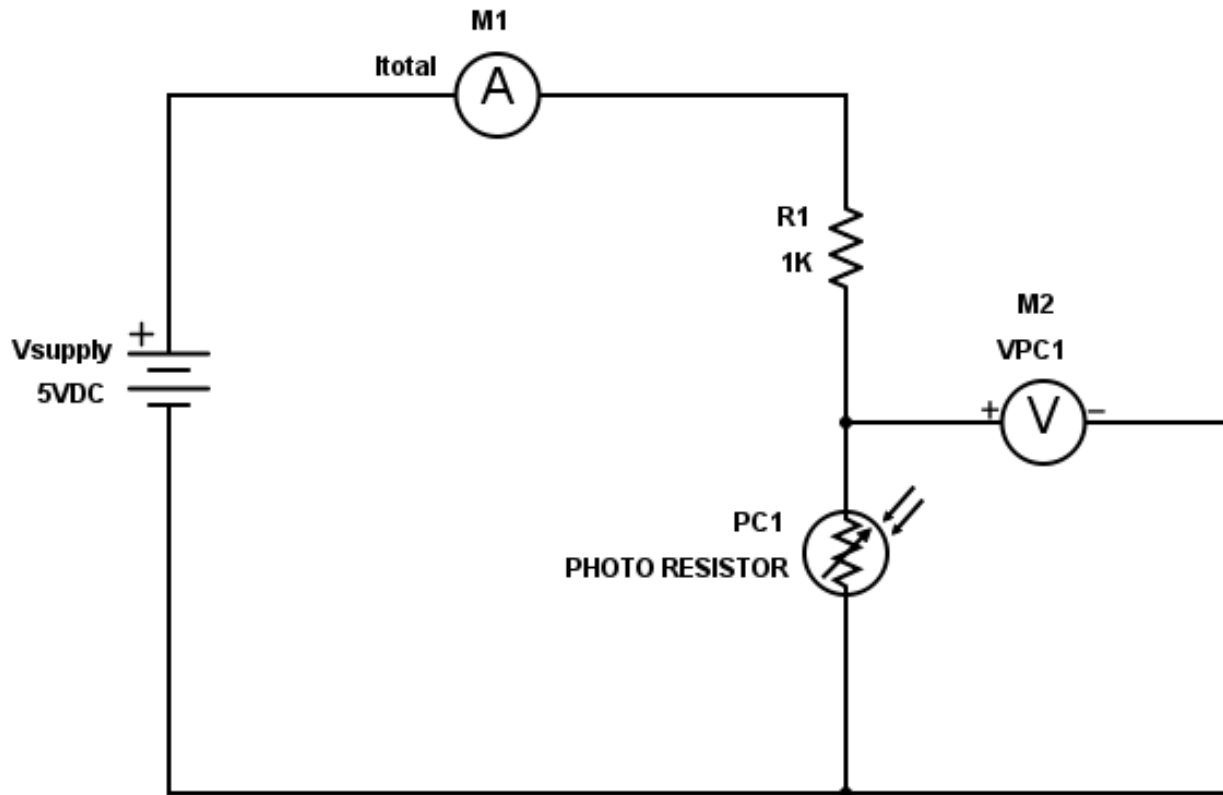
## Lab Project Objectives:

- a. Learn how to build an electronic sensor in TinkerCad Circuits.
- b. Learn how to create Common Separated Values file.
- c. Learn how to build a Predictive Model using Orange.

# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



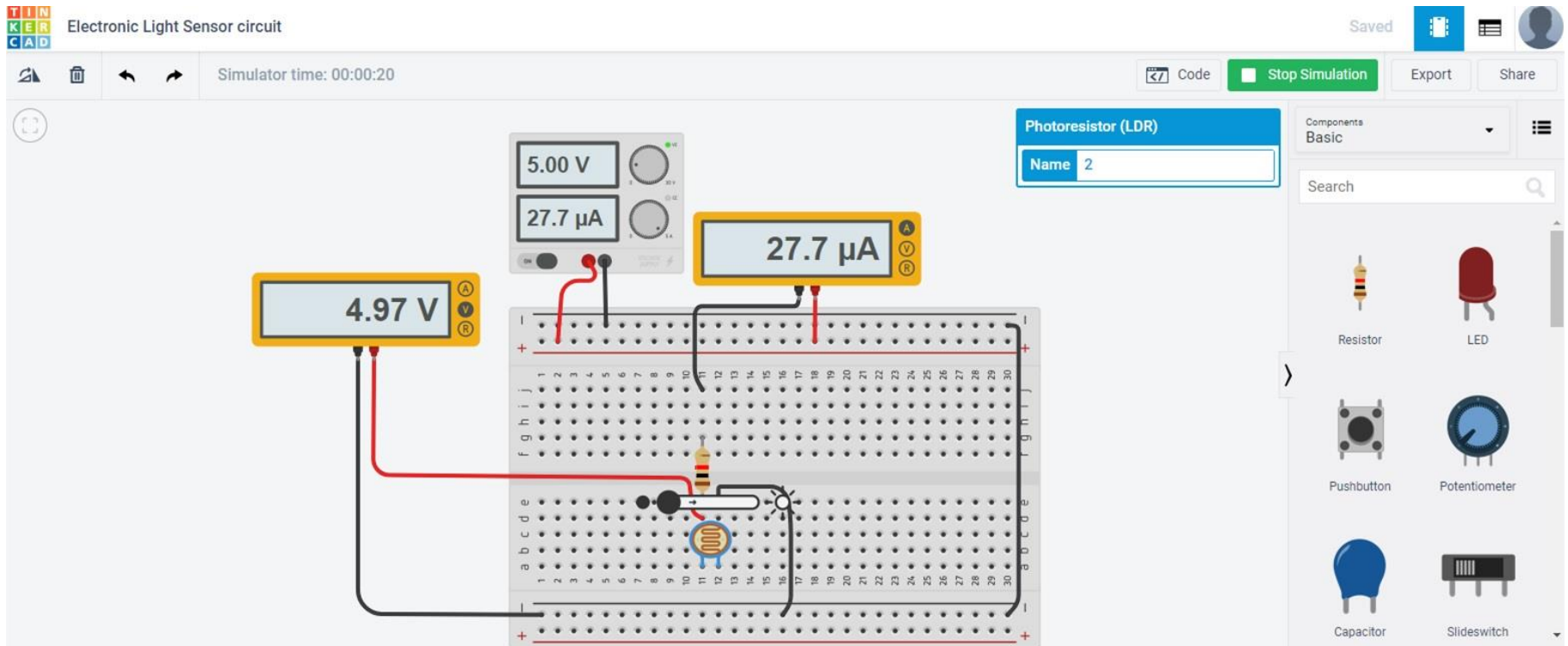
## Electronic Light Sensor Circuit.



# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



## TinkerCad Circuits: Electronic Light Sensor Circuit



Circuit Link:

<https://www.tinkercad.com/things/aUlgFjHDw5-electronic-light-sensor-circuit/editel>



# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



## Electronic Light Sensor Circuit Data(9 datapoints)

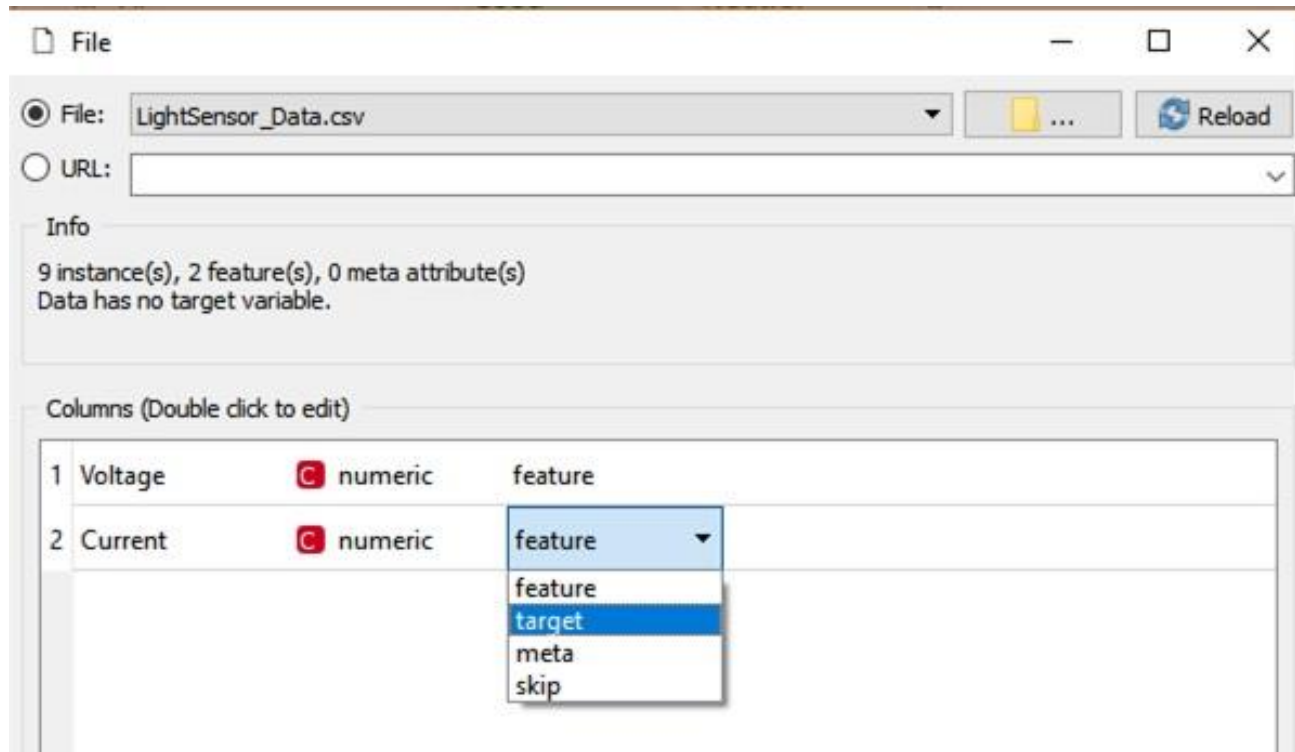
```
Voltage,Current
4.97,2.77E-05
4.42,5.78E-04
4.43,7.74E-04
4.06,9.45E-04
3.90,1.10E-03
3.64,1.36E-03
3.53,1.47E-03
3.32,1.68E-03
2.92,2.21E-03
```

**CSV file  
created in  
notepad**

# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



## Establishing an Electronic Light Sensor Circuit target variable.



# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



Data Table

Info

- 9 instances (no missing values)
- 1 feature (no missing values)
- Continuous target variable (no missing values)
- No meta attributes

Variables

- Show variable labels (if present)
- Visualize numeric values
- Color by instance classes

Selection

- Select full rows

Restore Original Order

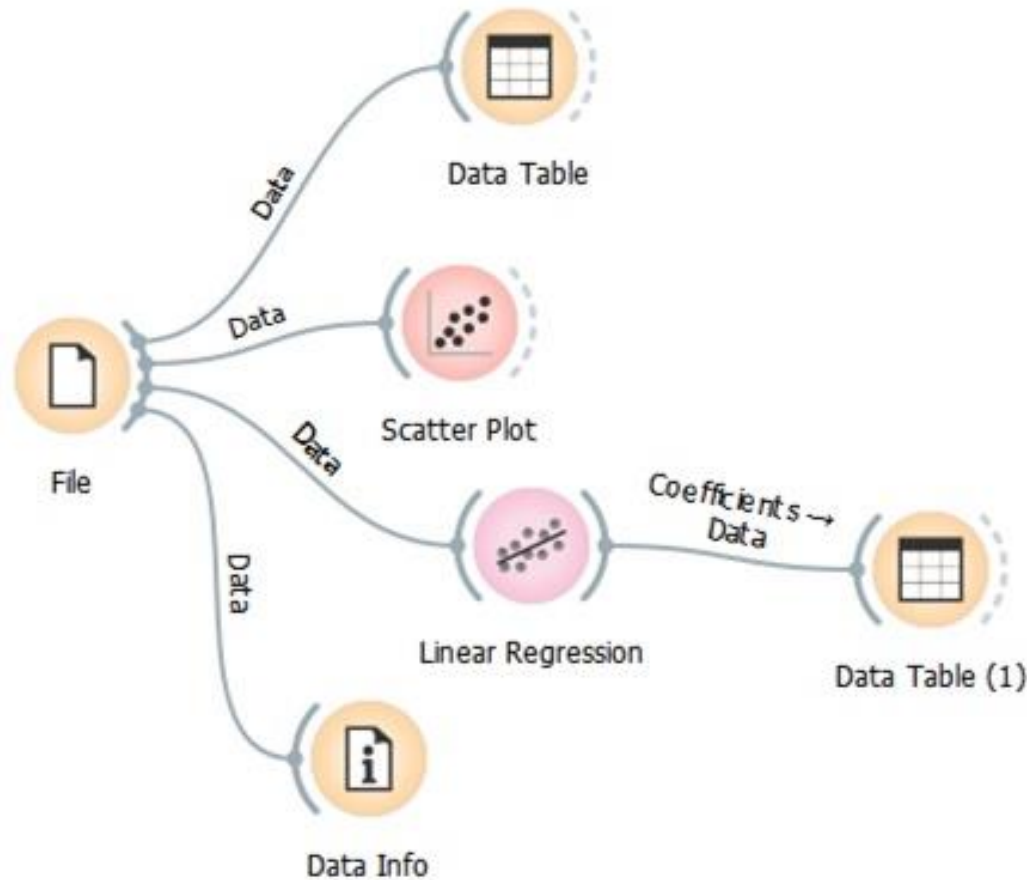
Send Automatically

	Current	Voltage
1	0.000028	4.97
2	0.000578	4.42
3	0.000774	4.43
4	0.000945	4.06
5	0.001100	3.90
6	0.001360	3.64
7	0.001470	3.53
8	0.001680	3.32
9	0.002210	2.92

# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



## Electronic Light Sensor Circuit Predictive Model Workflow

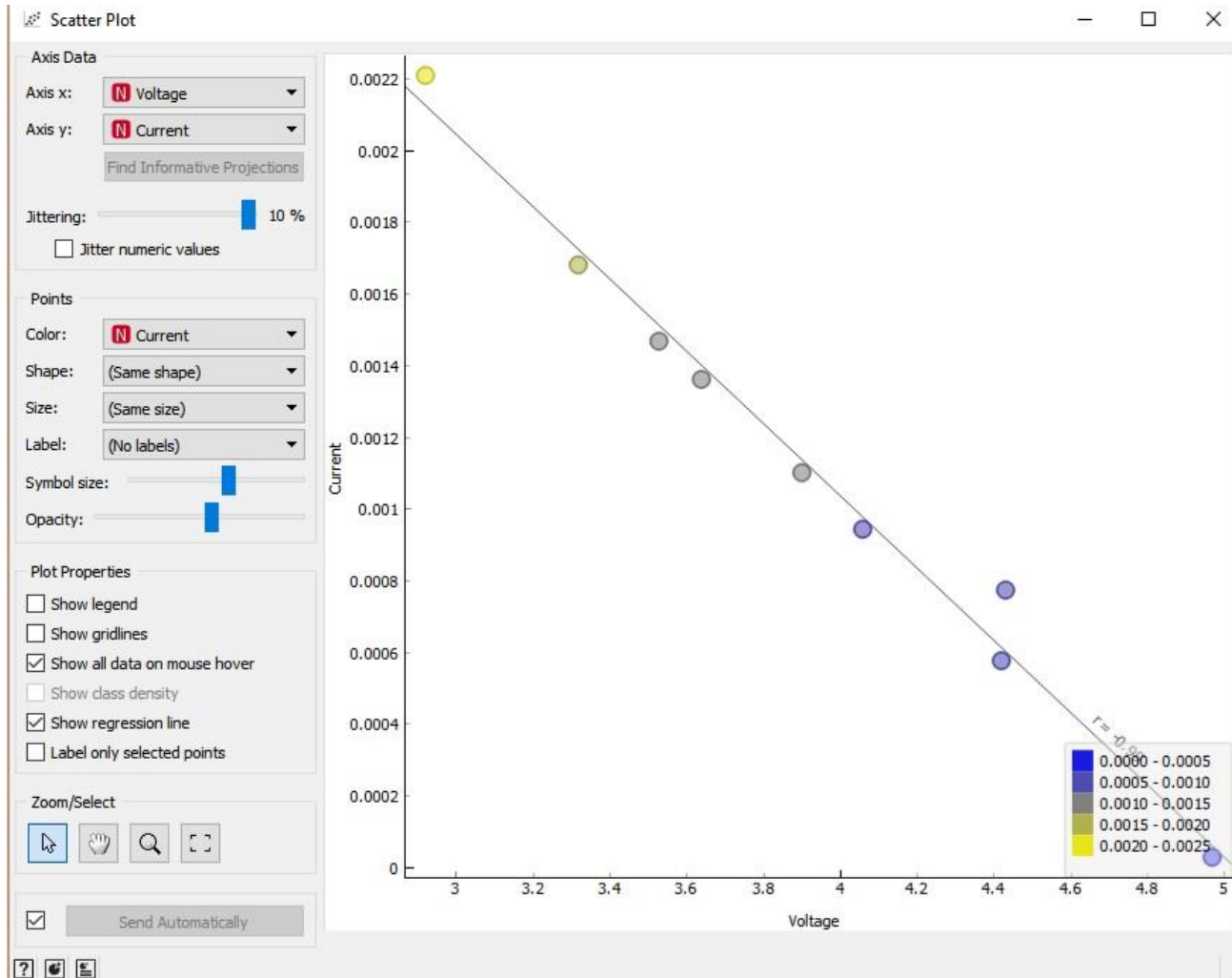


# Question 6



**Using slide 34: In configuring the input file, what feature is used as the target for the machine learning model in Orange?**

# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



Data Table (1)

Info  
2 instances (no missing values)  
1 feature (no missing values)  
No target variable.  
1 meta attribute (no missing values)

Variables  
 Show variable labels (if present)  
 Visualize continuous values  
 Color by instance classes

Selection  
 Select full rows

	name	coef
1	intercept	0.0050679
2	Voltage	-0.0010079

**Linear Regression equation.**

$$y = -0.0010079x + 0.0050679$$

Restore Original Order  
Report  
 Send Automatically

# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



**Predicting a new current value with a electronic sensor output voltage example.**

**Given:**

$$x = 1.68V$$

$$y = -0.0010079x + 0.0050679$$

$$y = -0.0010079(1.68) + 0.0050679$$

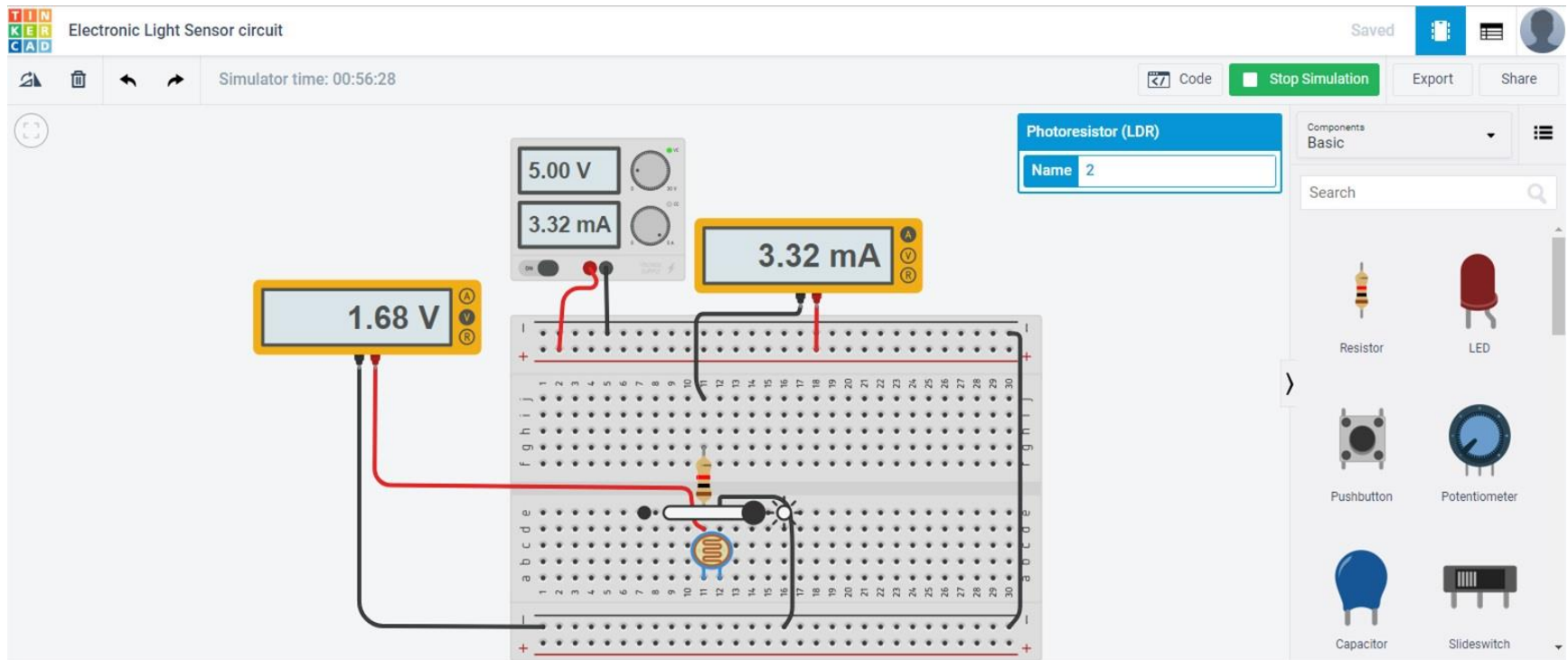
$$y = 0.003374628A = 3.374628mA$$



# Predicting the Electrical Behavior of a Light Sensor Circuit. . .



**Predictive output (Current) measured on the electronic light sensor model.**



# Data Collection with a Raspberry Pi. . .



Running (execute) the write file application in Python:

## Console Results

```
Python 3.5.3 Shell
File Edit Shell Debug Options Window Help
Python 3.5.3 (default, Jan 19 2017, 14:11:04)
[GCC 6.3.0 20170124] on linux
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: /home/pi/write_csv.py =====
>>> |
```

# Question 7



**What is the name of the online circuit modeling tool for constructing the Electronic Light Sensor Circuit?**

# Data Collection with a Raspberry Pi. . .



## Contents of person.csv.txt

```
/home/pi/person.csv.txt - pi@192.168.7.76 - Editor - WinSCP  
Name of Person, Age  
Peter, 22  
Jasmine, 21  
Sam, 24  
Don, 32
```