

# Examining Relationships in Data with Scatter Plots

---



**YK Sugishita**

SOFTWARE DEVELOPER / DATA SCIENTIST

[www.csdojo.io](http://www.csdojo.io)

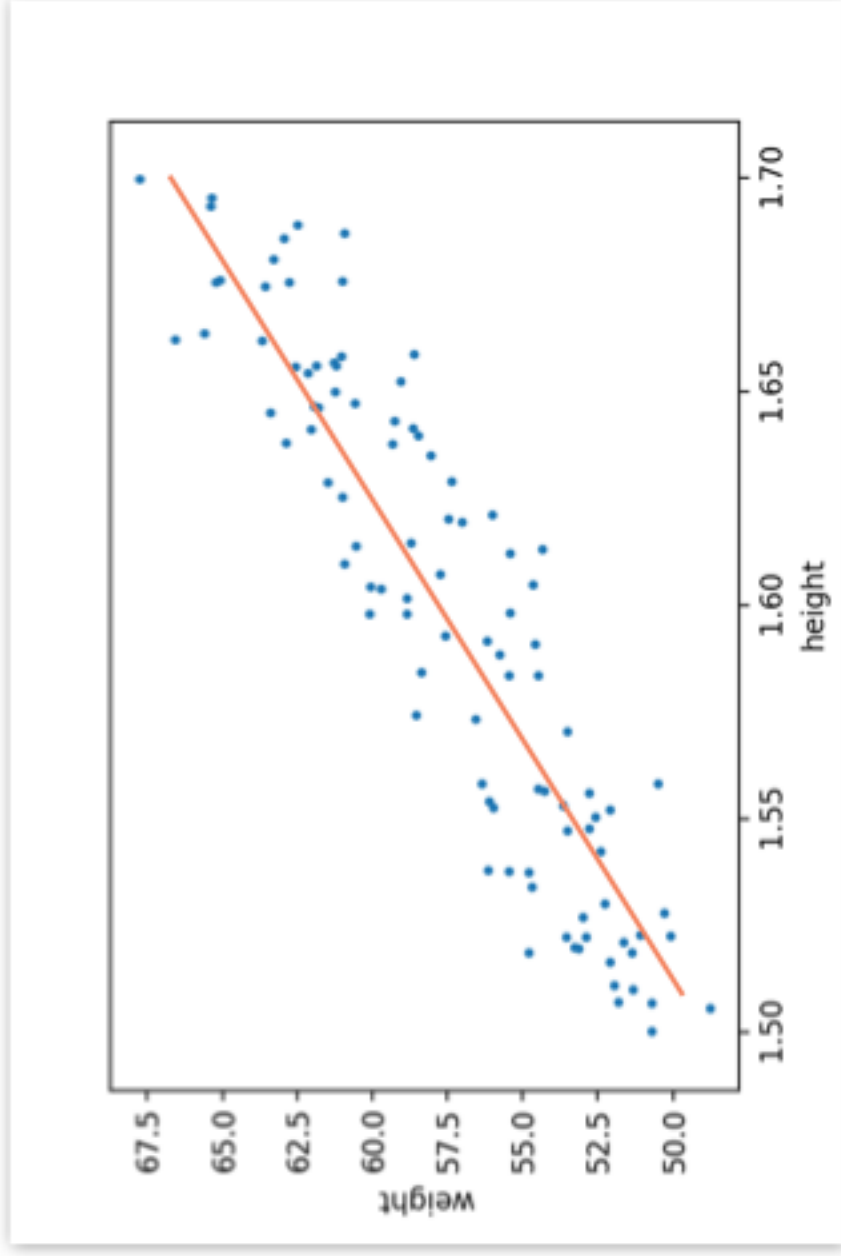


# Scatter Plots

A way to visualize how two numeric variables are related in your data

# Scatter Plots

# Scatter Plots



area	area	area	height	area
------	------	------	--------	------

## When to Use Scatter Plots

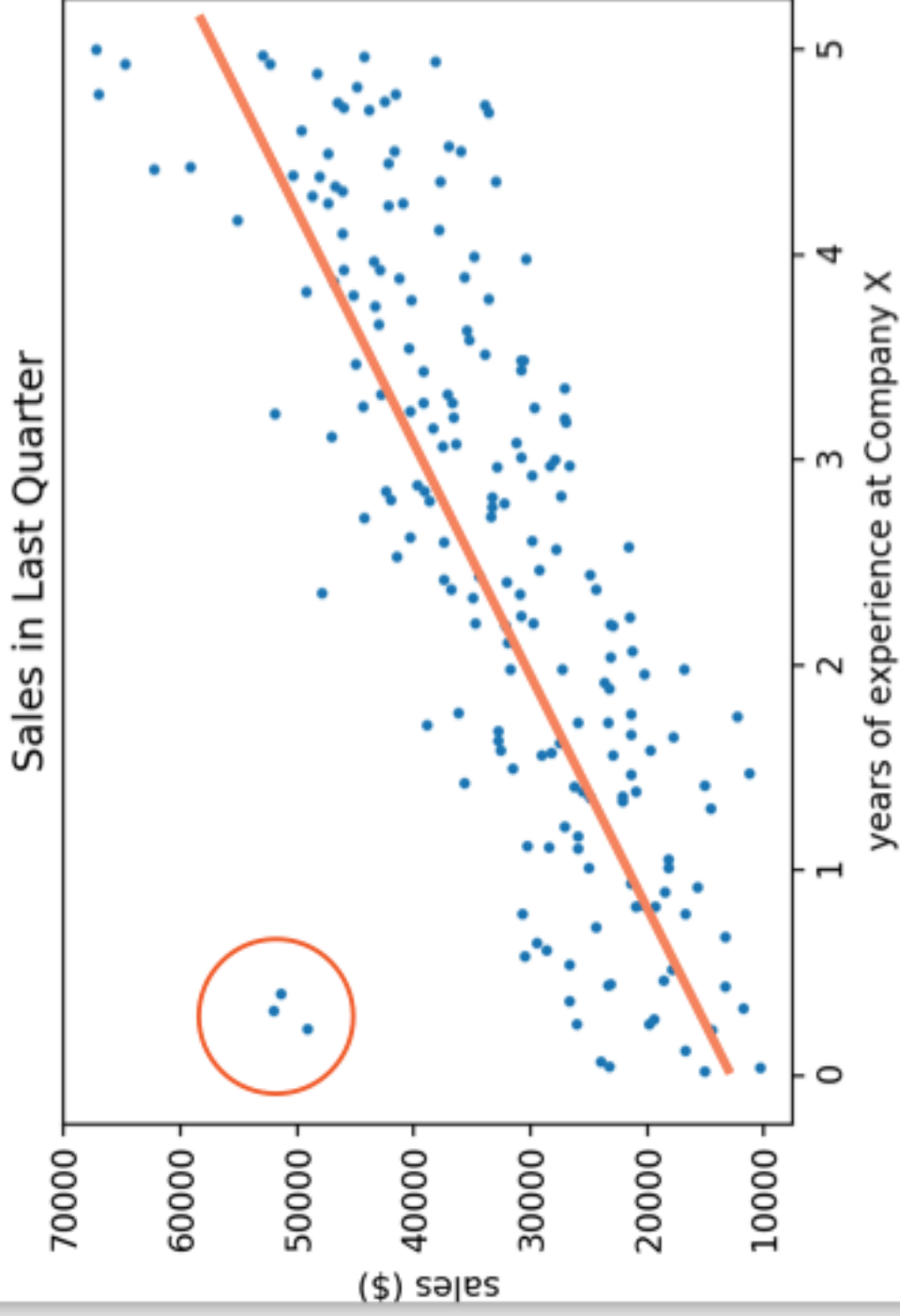
---



## Sales in Last Quarter

70000





## Why Scatter Plots

# Why Scatter Plots



Relationship between multiple variables

Helps you find outliers

# Creating Scatter Plots with Matplotlib

---



# Problem: Find Relationship Between GDP Per Capita and Life Expectancy

countries.csv

country	continent	year	lifeExpectancy	population	gdpPerCapita
---------	-----------	------	----------------	------------	--------------

## countries.csv

country	continent	year	lifeExpectancy	population	gdpPerCapita
Afghanistan	Asia	1952	28.801	8425333	779.4453145
Afghanistan	Asia	1957	30.332	9240934	820.8530296
Afghanistan	Asia	1962	31.997	10267083	853.10071
...					
Zimbabwe	Africa	1997	46.809	11404948	792.4499603
Zimbabwe	Africa	2002	39.989	11926563	672.0386227
Zimbabwe	Africa	2007	43.487	12311143	469.7092981

# Demo

## Creating Scatter Plots with Matplotlib

- Create scatter plots with `scatter()`
- Numpy's `log10()` function
- ***for item in ...*** syntax

## Practice Problem 3 - Scatter Plots

---

# Problem: Examine Relationship Between GDP (NOT per capita) and Life Expectancy in 2007

countries.csv

country	continent	year	lifeExpectancy	population	gdpPerCapita
Afghanistan	Asia	1952	28.801	8425333	779.4453145

## countries.csv

country	continent	year	lifeExpectancy	population	gdpPerCapita
Afghanistan	Asia	1952	28.801	8425333	779.4453145
Afghanistan	Asia	1957	30.332	9240934	820.8530296
Afghanistan	Asia	1962	31.997	10267083	853.10071
• • •					
Zimbabwe	Africa	1997	46.809	11404948	792.4499603
Zimbabwe	Africa	2002	39.989	11926563	672.0386227
Zimbabwe	Africa	2007	43.487	12311143	469.7092981

# Examine Relationship Between GDP and Life Expectancy in 2007

- Find GDP with `data.gdpPerCapita * data.population`
- Make a scatter plot
- Use both raw GDP and log of GDP
- Pause video

Demo

Example solution