



Chapter 5: Storytelling with Data Történetmesélés adatokkal



Big Data & Analytics

Cisco | Networking Academy®
Mind Wide Open™



Chapter 5 - Sections & Objectives

- 5.1 Adattörténet építése
 - Az adatokból történő érvelés alapjainak ismertetése.
- 5.2 A vizualizáció ereje
 - Elmagyarázza, hogyan lehet Python könyvtárakat használni a megfelelő vizualizációk létrehozásához kommunikációs célokra..
- 5.3 Felkészülés az 5. fejezet laboratóriumaira
 - A Big Data forrásainak leírása.

5.1 Adattörténet építése

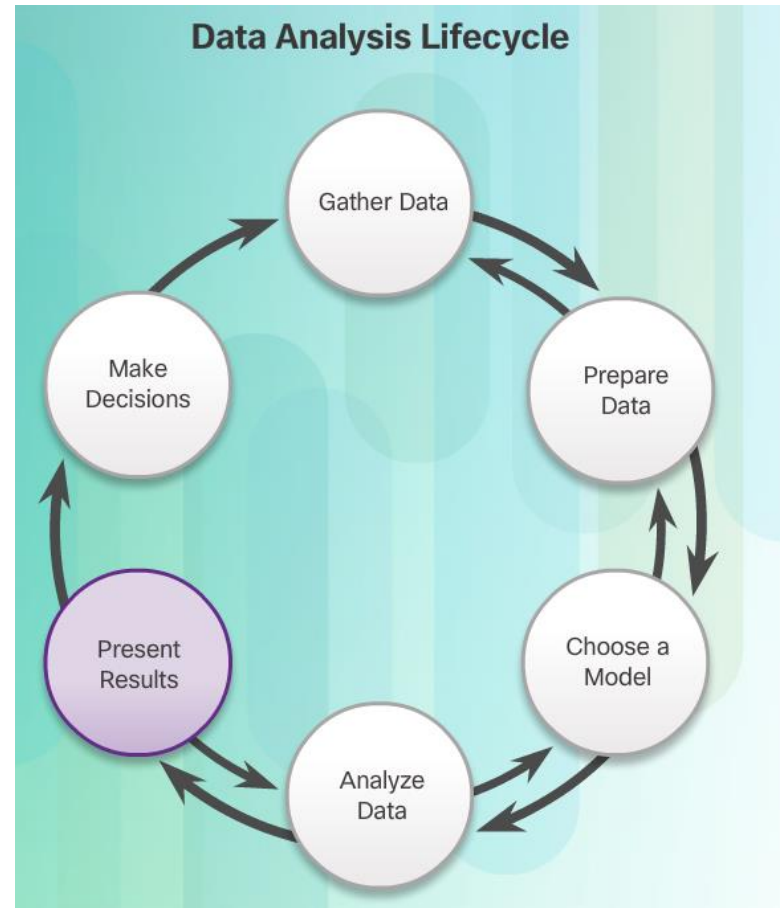




Building a Data Story

Telling a Story

- Az adatelemzés eredményei az adatelemzési életciklus Eredmények bemutatása részében jelennek meg.
- Az eredmények irányítják a döntéshozók által végrehajtott változtatásokat.
- Ne töltsön túl sok időt az adatokkal. Adjon eleget ahhoz, hogy megmagyarázza az álláspontját..





Building a Data Story

Audience

- Ki az Ön közönsége?
 - Ki fogja hallani a történetet?
 - Mi a hallgató motivációja?
 - Milyen szinten ismeri és ismeri a hallgató az üzleti problémát?
 - Mik a lehetséges reakciók?
- Hol van a közönség?
 - Online
 - Audio only - Csak hang
 - Face-to-face Személyes találkozás
 - Megosztják-e a prezentációt?
- Mikor érhető el a közönség?
 - Mi a teendő, ha valaki nem tud részt venni
 - A prezentáció rögzítése?
 - Titoktartási/biztonsági aggályok





Building a Data Story

Business Value and Goal

- Az üzleti érték különböző dolgokat jelent a különböző célközönségek számára, ezért legyen világos, hogy miért érdekel valakit az elmesélt történet.
- Mit szeretnél, hogy a közönség tagjai mit vigyenek magukkal?
- Mi a cselekvésre való felhívás, ha van ilyen?





Building a Data Story

Bizonyítékok felhasználása

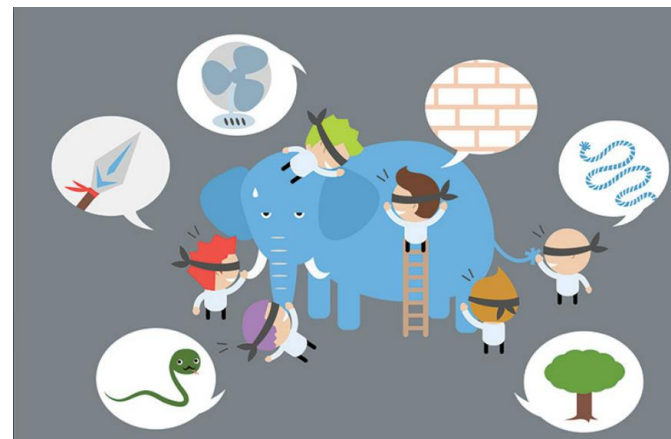
- A végcél szempontjából kritikusnak kell Lennie
- Ha egy bizonyíték nem támasztja alá a következtetéseket, vagy másodlagos az elsődleges témához képest, fontolja meg, hogy kihagyja a bizonyítékot a prezentációból.



Building a Data Story

Deduktív érvelés

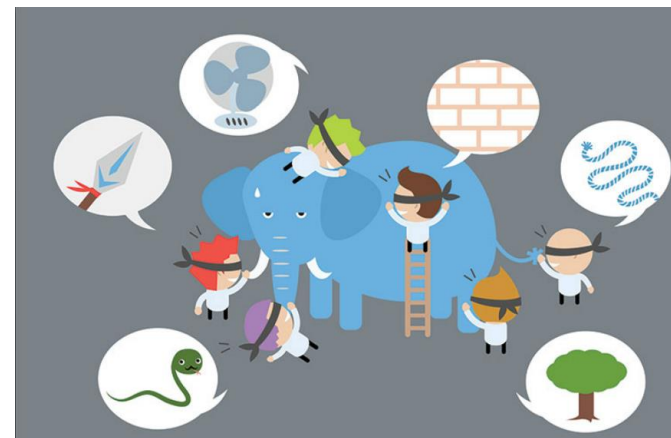
- Tények vagy premisszák felhasználása következtetés levonásához
- "Top-down"-nak tekinthető, mert egy általános előfeltevéstől halad az általános előfeltevésből levezetett konkrét tények felé.
- A helyes deduktív érvelés igaz következtetésekhez vezet.
- Példa: szillogizmus Minden emlősnek van szeme. Az emberek emlősök. Ezért az embereknek van szemük.



Building a Data Story

Inductive Reasoning

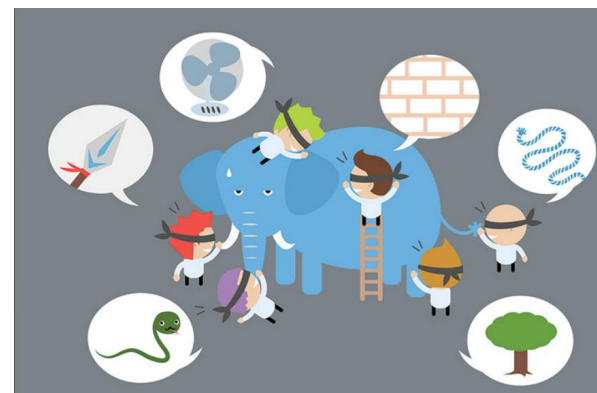
- A specifikusról az általánosra való áttérés
- Hozzon létre egy következtetést a megfigyelések, minták és hipotézisek alapján.
- Mintát veszünk egy populációból, megvizsgáljuk a mintát, majd olyan következtetéseket vonunk le, amelyekről azt hisszük, hogy a teljes populációra igazak lesznek.
- Győződjön meg arról, hogy a minta reprezentálja azt a populációt, amelyre a következtetést alkalmazzák



Building a Data Story

Fallacies Tévedések

- Az érvelés nem feltétlenül alkalmazza a logika szabályait
- Az érvelés kihagyhat vagy félreértelmezhet egy kulcsfontosságú előfeltevést.
- A következtetés nem feltétlenül következik logikusan a premisszából.(s)
- Formai tévedés
 - If milk is kept in the refrigerator, it will not spoil. The milk is spoiled. Therefore, the milk was not kept in the refrigerator.
- Informális tévedés
 - Some people have psychic powers. Can you prove it? No one has been able to disprove it.





5.2 The Power of Visualization



Cisco | Networking Academy®
Mind Wide Open™

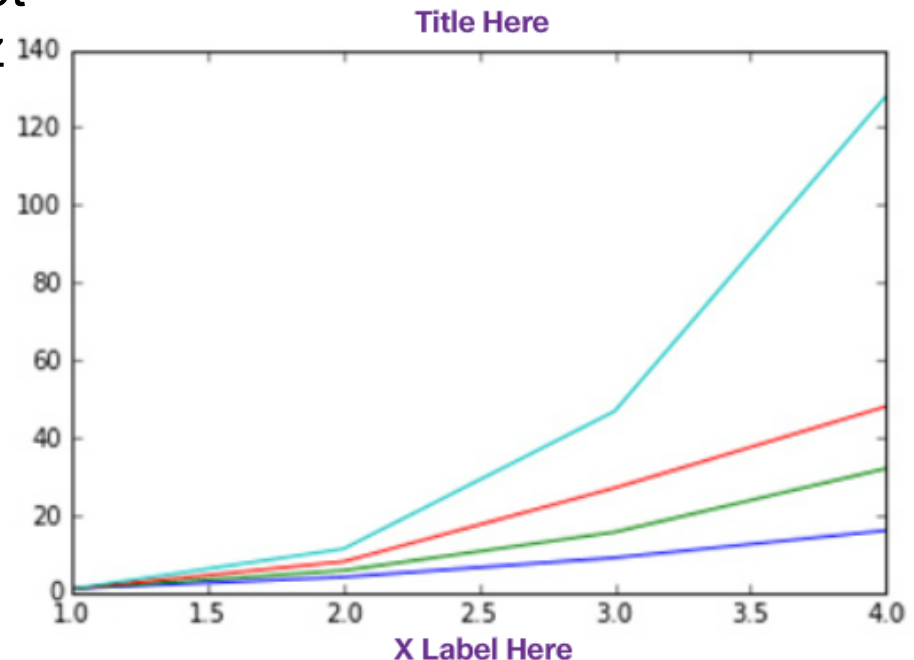


The Power of Visualization

Pyplot

- Pyplot is a Matplotlib module.
- A Pyplot tartalmaz egy gyűjteményt a stílusfüggvényekből, amelyeket a plot létrehozásához és testreszabásához használhat.

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.plot([1,2,3,4], [1, 4, 9, 16])
plt.plot([1,2,3,4], [1, 5.7, 15.6, 32])
plt.plot([1,2,3,4], [1, 8, 27, 48])
plt.plot([1,2,3,4], [1, 11.3, 46.8, 128])
plt.xlabel('X Label Here')
plt.ylabel('Y Label Here')
plt.title('Title Here')
plt.show()
```





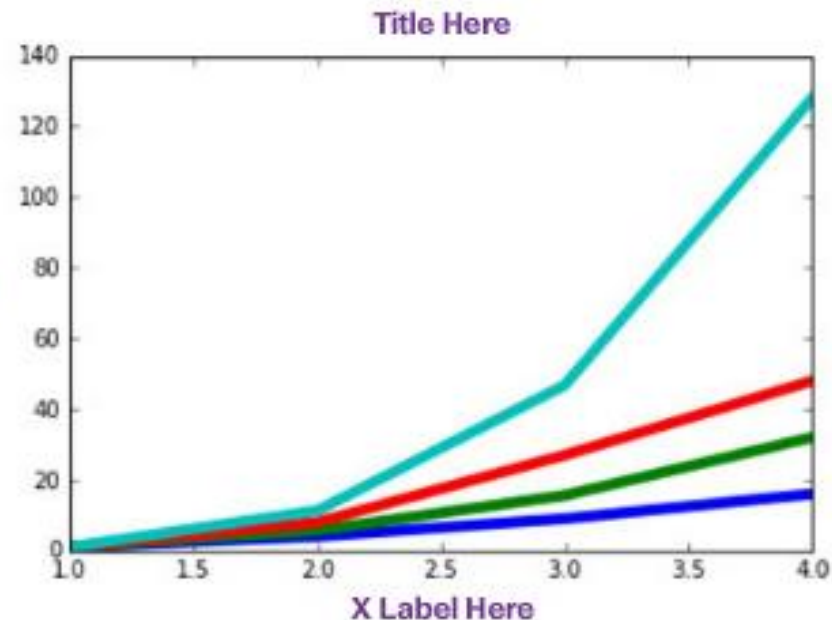
The Power of Visualization

Pyplot Custom Style Sheet

- In pyplot you can create a custom style sheet so all plots have the same style feature and you avoid making minor errors to the inline code.
- If you store the style sheet in a non-default location, you must provide path information when you reference it.

```
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline
plt.style.use('mystyle.mplstyle')
plt.plot([1,2,3,4], [1, 4, 9, 16])
plt.plot([1,2,3,4], [1, 5.7, 15.6, 32])
plt.plot([1,2,3,4], [1, 8, 27, 48])
plt.plot([1,2,3,4], [1, 11.3, 46.8, 128])
plt.xlabel('X Label Here')
plt.ylabel('Y Label Here')
plt.title('Title Here')
plt.show()
```

Y Label Here

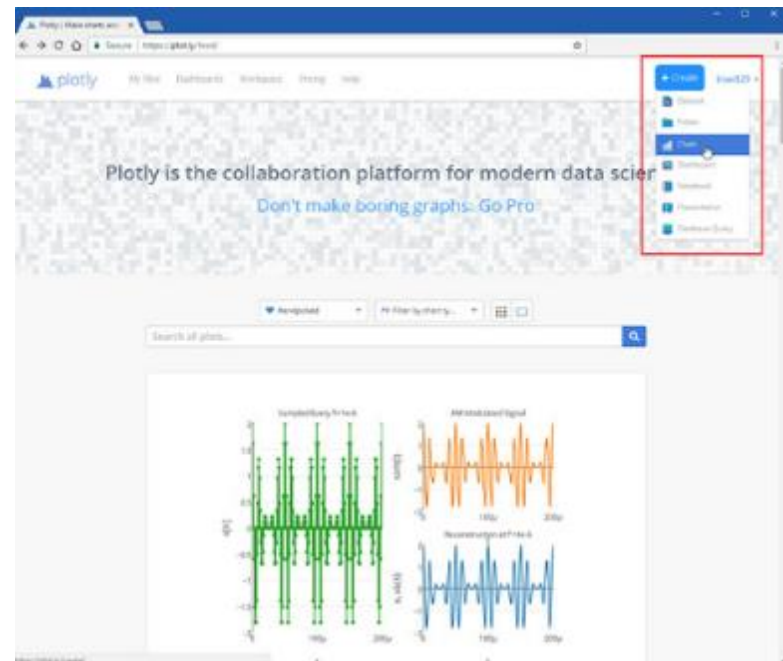
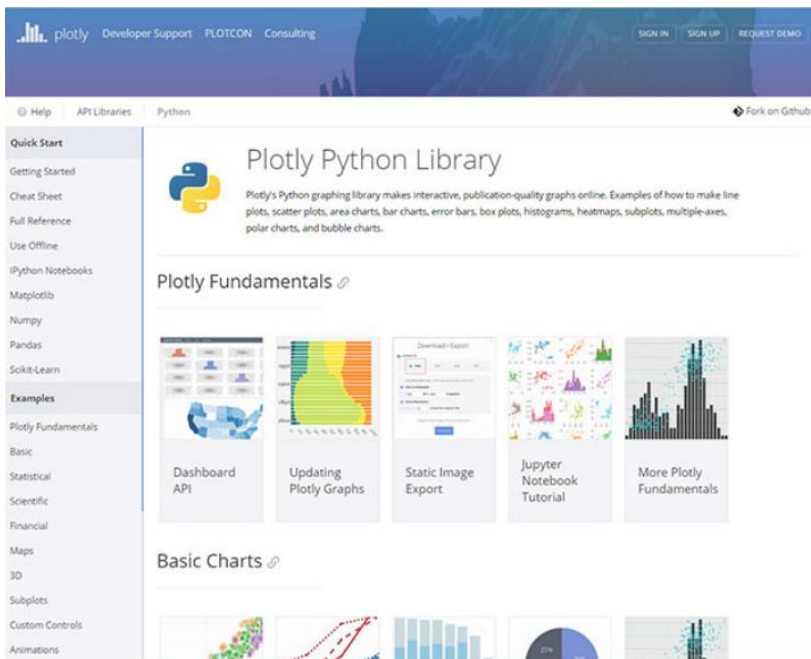




The Power of Visualization

Plotly

- A Plotly egy online eszköz adatvizualizációk létrehozására..
- Források, többek között ingyenes tartalmak, API-könyvtárak, számkonvertálók, alkalmazások a Google Chrome-hoz és egy nyílt forráskódú JavaScript-könyvtár.
- Kód, képek és adatok exportálhatók

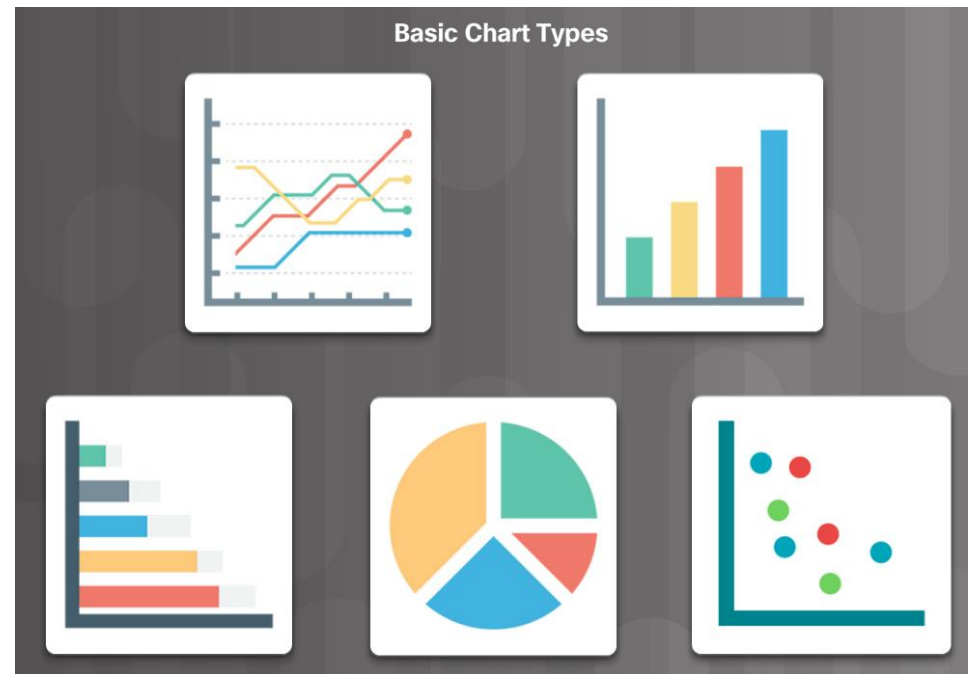




The Power of Visualization

Common Types of Data Visualizations

- Hány változó?
- Hány adatpont van az egyes változóknak?
- Az adatok időbeli vagy tételek összehasonlítása?
- Legnépszerűbb diagramok: vonal-, oszlop-, oszlop-, oszlop-, kör- és szórásdiagramok

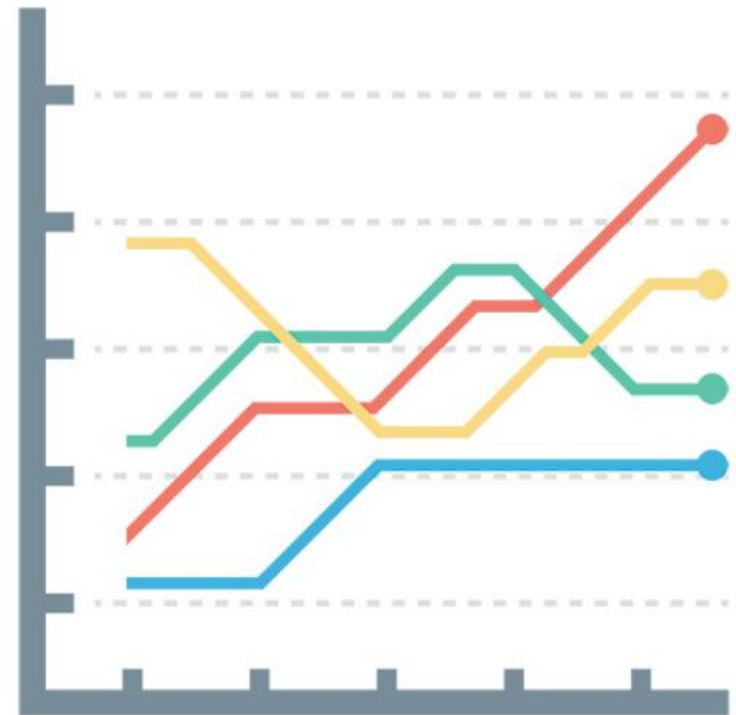




The Power of Visualization

Line Charts

- One of the most commonly used
 - Used when there is a continuous set of data, the number of data points is high, and you would like to show a trend in data over time
- Examples
 - Quarterly sales for past five years
 - Number of customers per week in the year
- Best practices
 - Label axes.
 - Plot time on the x-axis (horizontal).
 - Plot data values on the y-axis (vertical).
 - Keep data sets to a minimum.
 - Minimize gridlines.
 - Modify the axis starting point if necessary.





The Power of Visualization

Column Chart Oszlopdiagram

■ Positioned Vertically

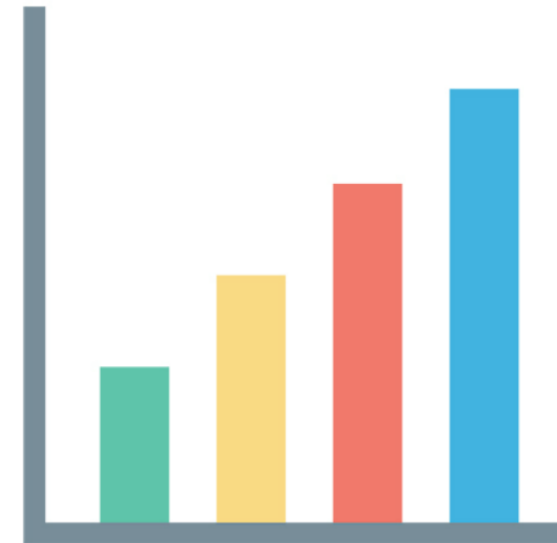
- Most commonly used when you want to display the value of a specific data point compared across similar categories

■ Examples

- Population of five nations
- Yearly sales for four companies

■ Best practices

- Label axes.
- If time is used, plot on the x-axis (horizontal).
- Use solid colors for columns.
- Avoid using more than 7 categories on the horizontal axis.
- Start the y-axis value at zero.
- Spacing between columns should be about half the width of a column.





The Power of Visualization

Bar Chart

■ Positioned Horizontally

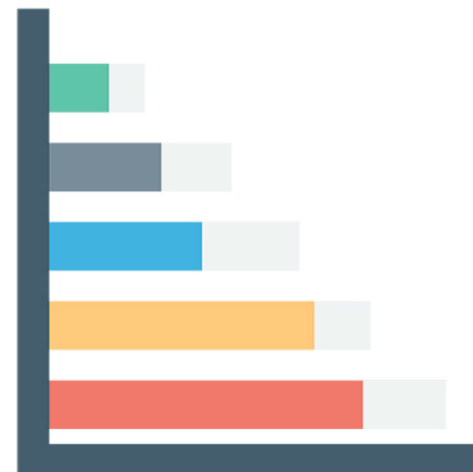
- Most commonly used when you want to display the value of a specific data point compared across similar categories and the names for each data point is long.

■ Examples

- GDP of 25 nations
- Car sales by salesman

■ Best practices

- Label axes.
- Order bars from longest to shortest.
- Use solid colors for bars.
- Avoid using more than 7 categories on the horizontal axis.
- Start the x-axis value at zero.
- Spacing between rows should be about half the width of a column.





The Power of Visualization

Pie Charts Kőrdiagramok

■ Pie chart

- Used to show the composition of a static number.
- Segments show a percentage of that number
- Segments total 100%

■ Examples

- Annual expenses by type
- Energy sources by type used

■ Best practices

- Keep categories to a minimum. Consolidate when necessary.
- Use different colors for different segments and order by size.
- Ensure segment values total 100%.

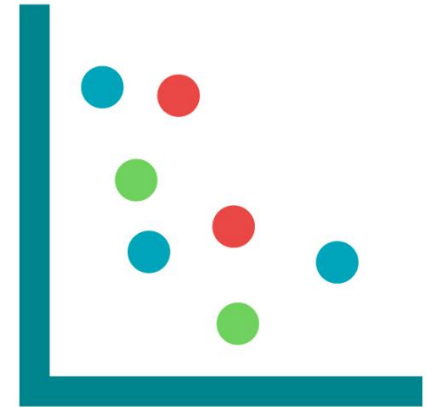




The Power of Visualization

Scatter Plot Szóródási ábra

- Clustering and Correlation Visualizations
 - Used to show correlation or distribution of data points
 - Useful in showing clustering or identifying data outliers
- Examples
 - Comparing life expectancy to GDP
 - Comparing daily sales of ice cream to average temperature
- Best practices
 - Label axes.
 - Ensure data set is large enough.
 - Start y-axis at zero. X-axis start value depends on data.
 - Consider adding a trend line, but don't use more than two.





5.3 Chapter 5 Labs



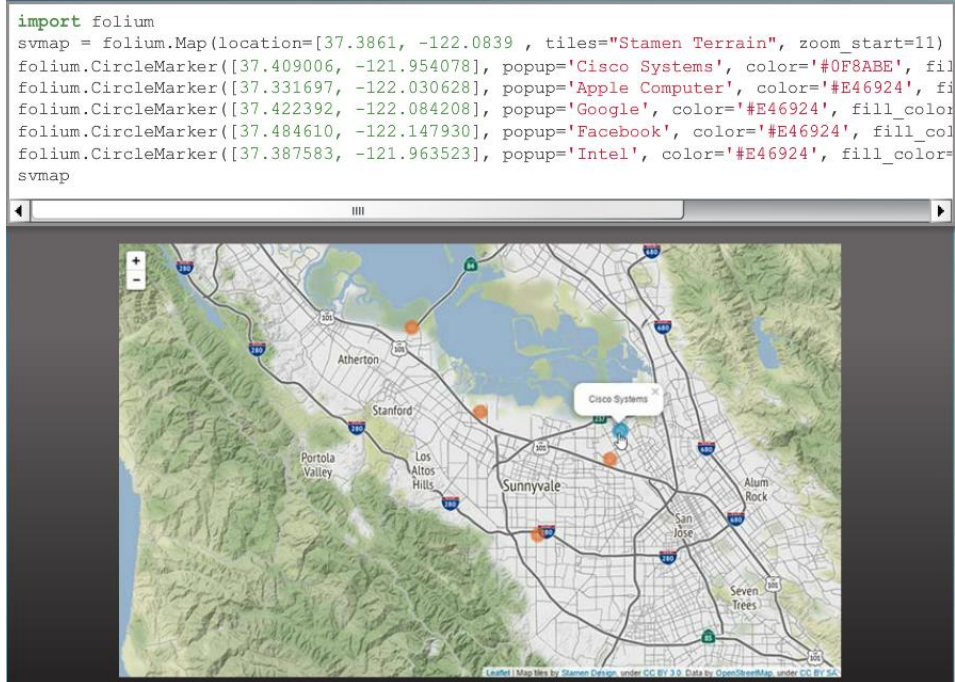
Cisco | Networking Academy®
Mind Wide Open™



Chapter 5 Preparation for Chapter 5 Labs

Folium Library

- Egyesíti a Python szkriptek erejét a Leaflet.js könyvtár leképezési képességeivel.
- Lehetővé teszi a Python adatkeretek megjelenítését egy interaktív Leaflet-térképben
- Csempekészlet - raszteres vagy vektoros adatok gyűjteménye, amely képes megjeleníteni egy térképet mobileszközökön és böngészőben.





5.4 Summary



Cisco | Networking Academy®
Mind Wide Open™



Chapter Summary

Summary

- Az adatokat vizualizációk segítségével lehet összefoglalni, hogy mások is megértsék az adatokat..
- Tudnia kell, hogy ki a célközönsége, hol van, és mikor érhető el.?
- A bemutatott bizonyítékok származhatnak deduktív vagy induktív érvelésből, és nem szenvedhetnek (formális vagy informális) logikai tévedésben..
- A deduktív érvelés tényeket, tételeket vagy más igazságot tartalmazó állításokat használ a következtetés levonásához..
- Az induktív érvelés megfigyelések, minták és hipotézisek alapján von le következtetést..
- A megjelenítésekben használt diagramtípusok a vonal-, oszlop-, oszlop-, oszlop-, kör- és szórásdiagramok.
- A Pyplot egy matplotlib kiterjesztés, amely stílusfüggvényeket tartalmaz a plot létrehozásához és testreszabásához..
- A Plotly egy online eszköz, amelyet vizualizáció létrehozására használnak.

