Data Visualization Systems in Python

Agenda

- Importance of Data Visualization
- Principles of Analytical Graphs
- Basic Graphic systems in python
- Elements of a Graph
- Storytelling with Graphs
- Benefits of data visualization

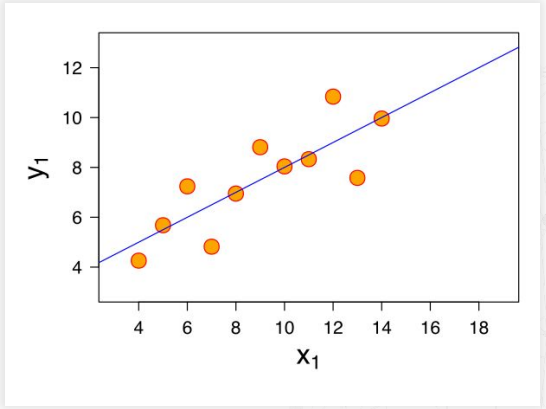
</talentlabs>

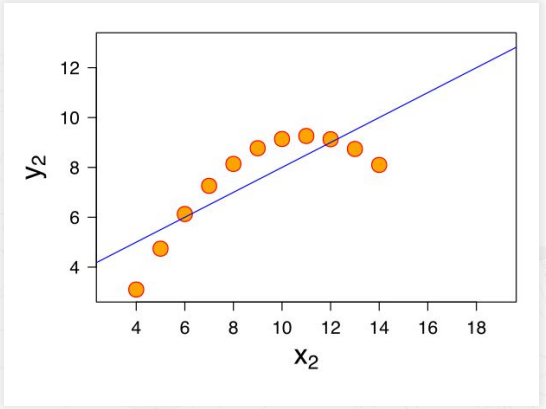
Data Visualization

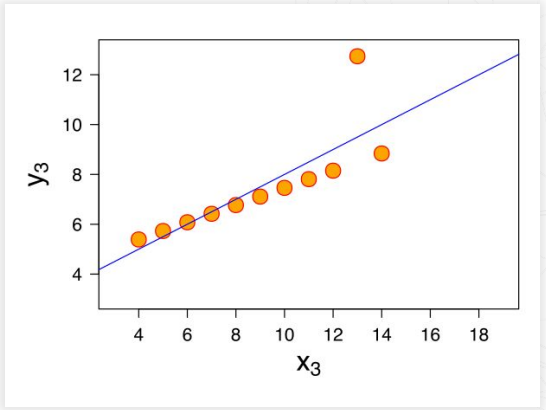
- Integral part of Exploratory Data Analysis
- Allows a data analyst to “look at” their data and deduce assumptions
- Get to know the variables and relationships between them.
- Analyzing reports helps business stakeholders focus on the areas that require attention
- Faster Decision Making
- Make sense of Complicated Data

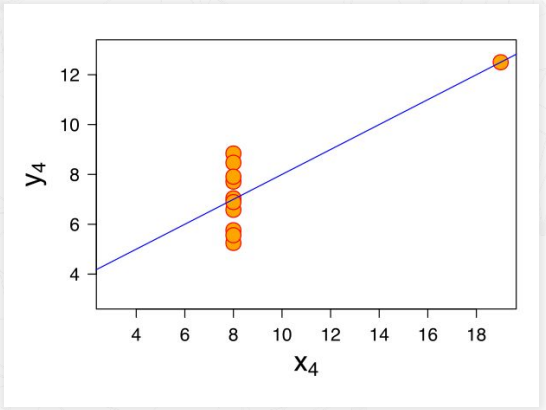
</talentlabs>

Why should we plot? - Anscombe's quartet









</talentlabs>

Principles of Analytical Graphs

1. Showing comparisons is really the basis of all good scientific investigation.
2. Show causality, mechanism, explanation, systematic structure
3. The real world is multivariate. For anything that you might study, there are usually many attributes that you can measure.
4. Analytical presentations ultimately stand or fall depending on the quality, relevance, and integrity of their content.
5. Integrate evidence
6. Data graphics should be appropriately documented with labels, scales, and sources. A general rule for me is that a data graphic should tell a complete story all by itself.

</talentlabs>

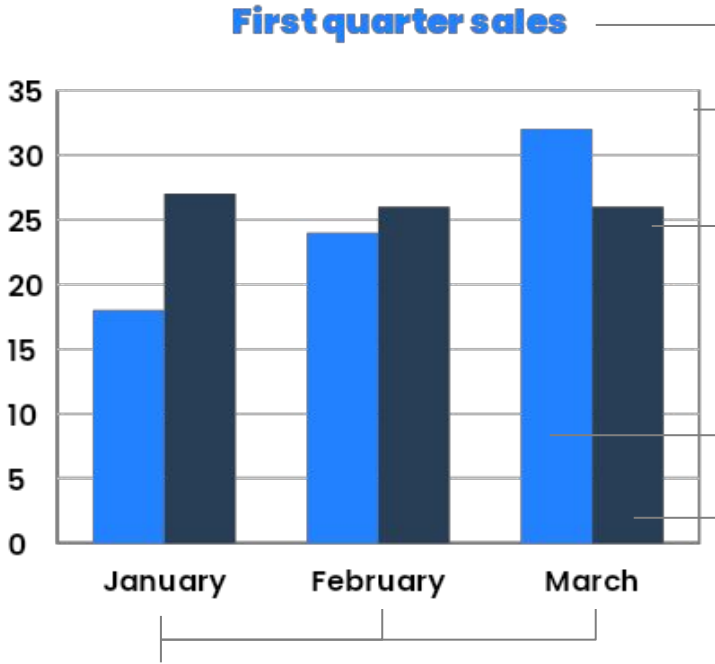
Graphing Systems and Tools in Python

</talentlabs>

Basic Graphic Tools in Python

Characteristics	Matplotlib	Seaborn
Use Cases	Matplotlib plots various graphs using Pandas and Numpy	Seaborn is the extended version of Matplotlib which uses Matplotlib along with Numpy and Pandas for plotting graphs
Complexity of Syntax	It uses comparatively complex and lengthy syntax	It uses comparatively simple syntax which is easier to learn and understand
Multiple figures	Matplotlib has multiple figures can be opened	Seaborn automates the creation of multiple figures which sometimes leads to out of memory issues
Flexibility	Matplotlib is highly customizable and powerful	Seaborn avoids a ton of boilerplate by providing default themes which are commonly used

Elements of a Graph



January, February, and March are categories

Title

Plot

Data point

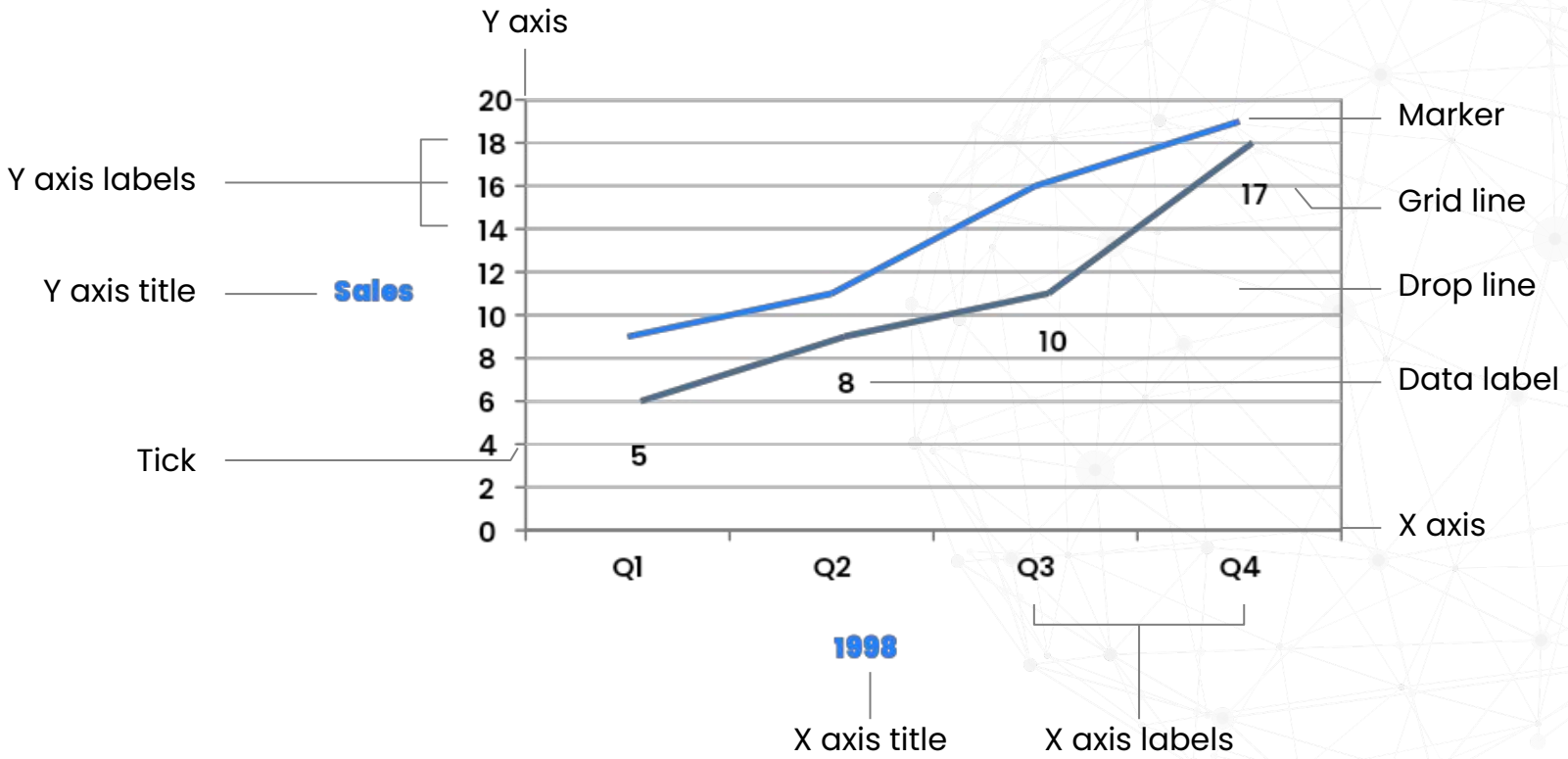
Legend

The three bars representing internet sales form a series

The darker bars representing in-store sales form another series

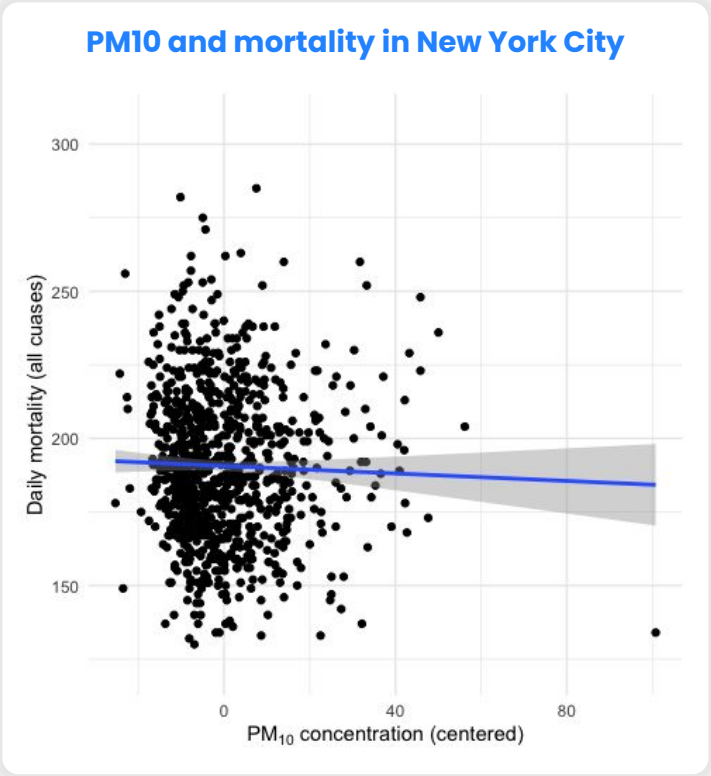
</talentlabs>

Elements of a Graph



</talentlabs>

Storytelling with Data – An Example



The Data

- Each point on the plot represents the average PM10 level for that day (measured in micrograms per cubic meter) and the number of deaths on that day.
- The PM10 data come from the U.S. Environmental Protection Agency and the mortality data come from the U.S. National Center for Health Statistics.

The Analysis

- This is a bivariate plot showing two variables in this dataset.
- From the plot it seems that there is a slight negative relationship between the two variables.

The Insight

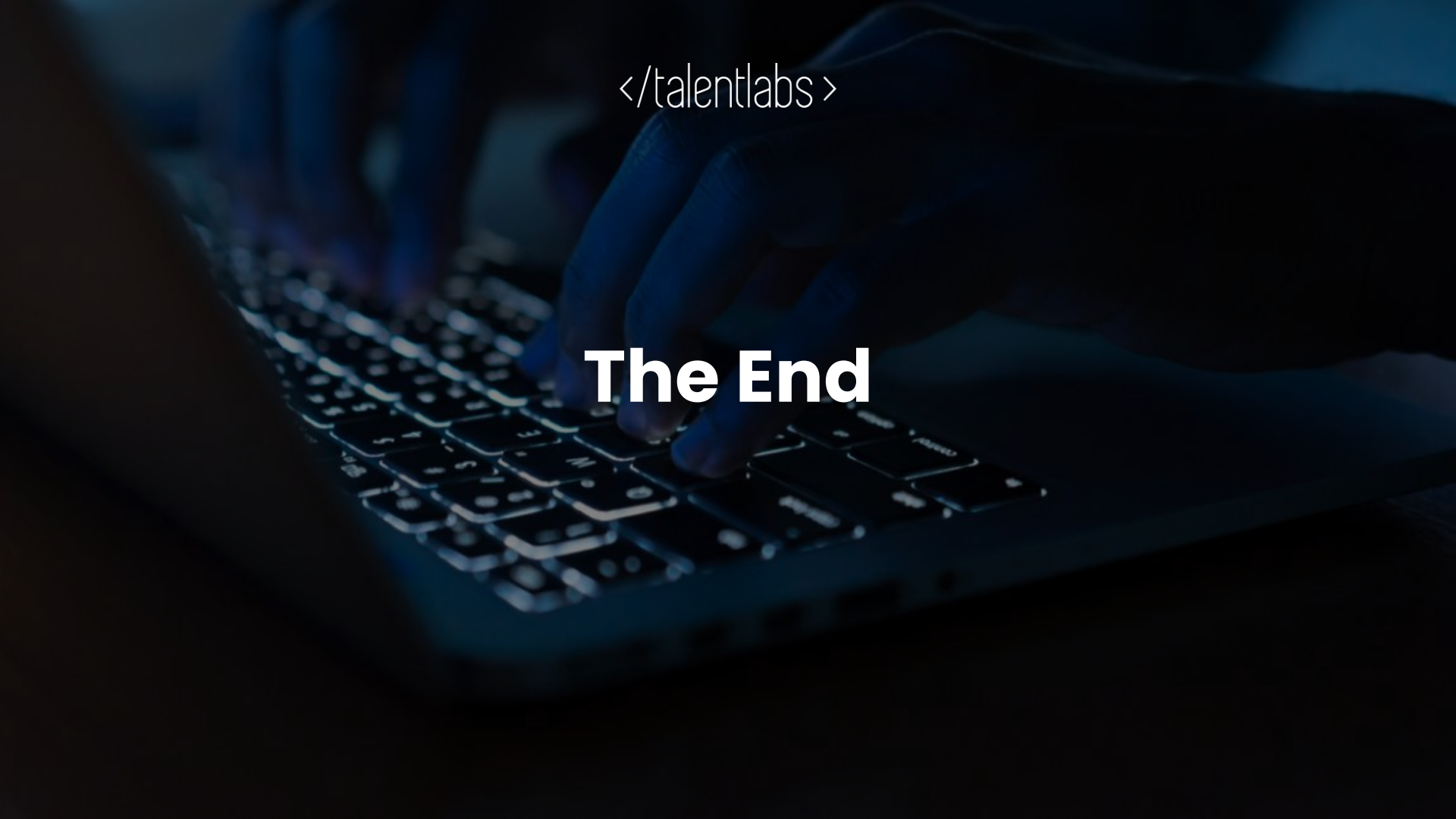
- That is, higher daily average levels of PM10 appear to be associated with lower levels of mortality (fewer deaths per day).

</talentlabs>

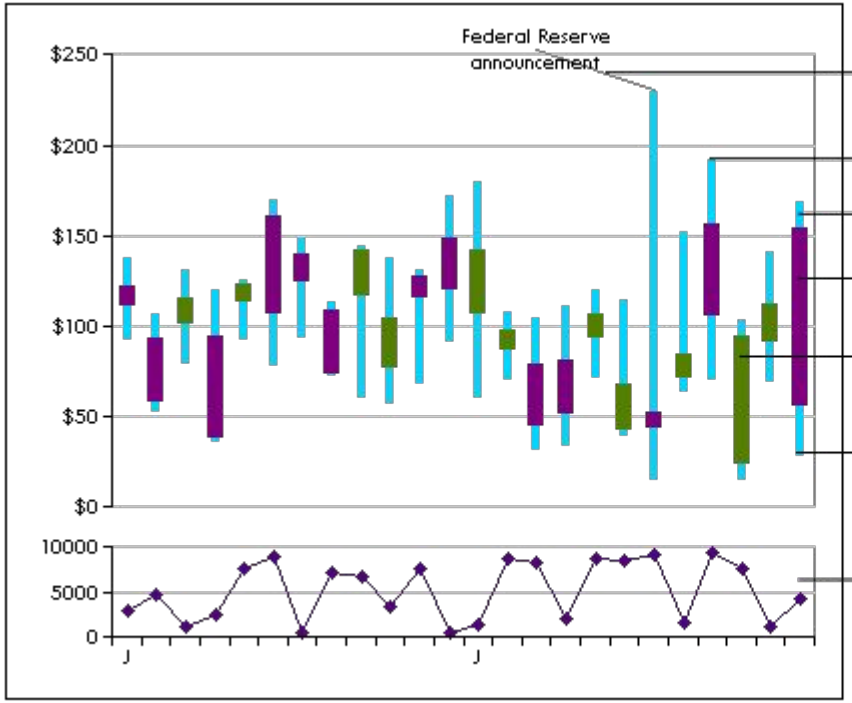
Benefits of data visualization

- Better monitoring
- Taking decisions
- Helps business intelligence
- Accurate data
- Finding efficiencies/inefficiencies

</talentlabs>


</talentlabs>
The End

Elements of a Graph



Label line

Highest price attained for that period

High-low line

Open-close bar showing gain

Open-close bar showing loss

Lowest price attained for that period

Study showing volume data

</talentlabs>

OLD SLIDES

Data Visualization Systems in Python

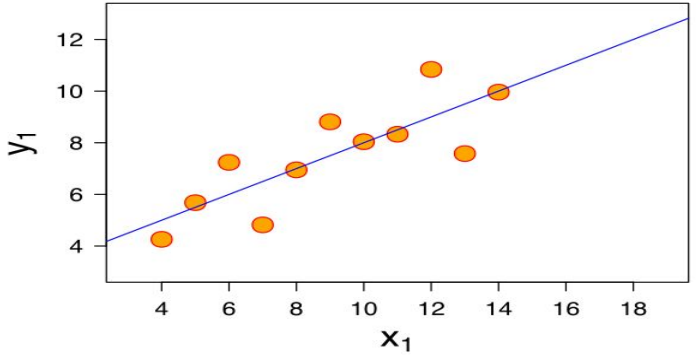
Agenda

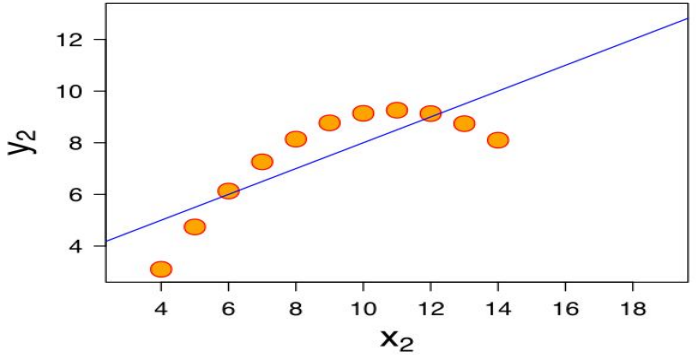
- Importance of Data Visualization
- Principles of Analytical Graphs
- Basic Graphic systems in python
- Elements of a Graph
- Storytelling with Graphs
- Advanced Plotting using Plotly (optional)

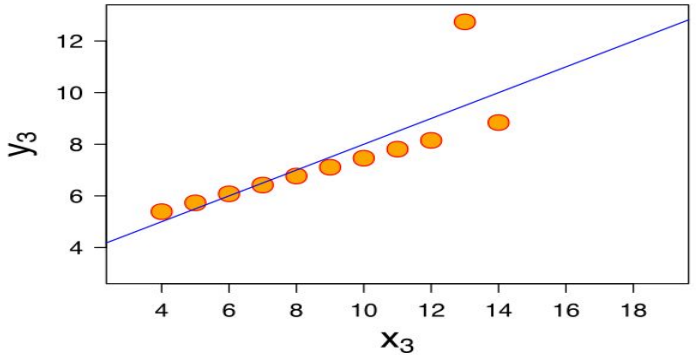
Data Visualization

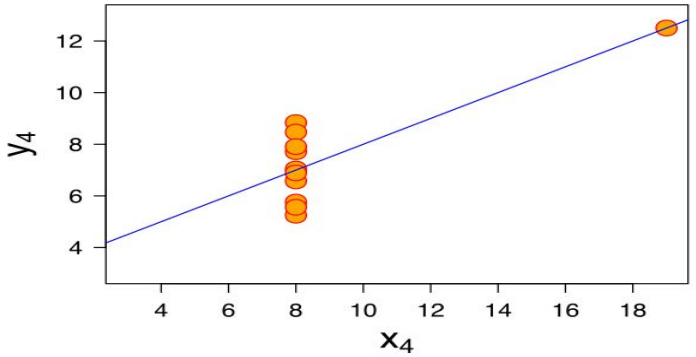
- Integral part of Exploratory Data Analysis
- Allows a data analyst to “look at” their data and deduce assumptions
- Get to know the variables and relationships between them.
- Analyzing reports helps business stakeholders focus on the areas that require attention
- Faster Decision Making
- Make sense of Complicated Data

Why should we plot? - Anscombe's quartet









Principles of Analytical Graphs

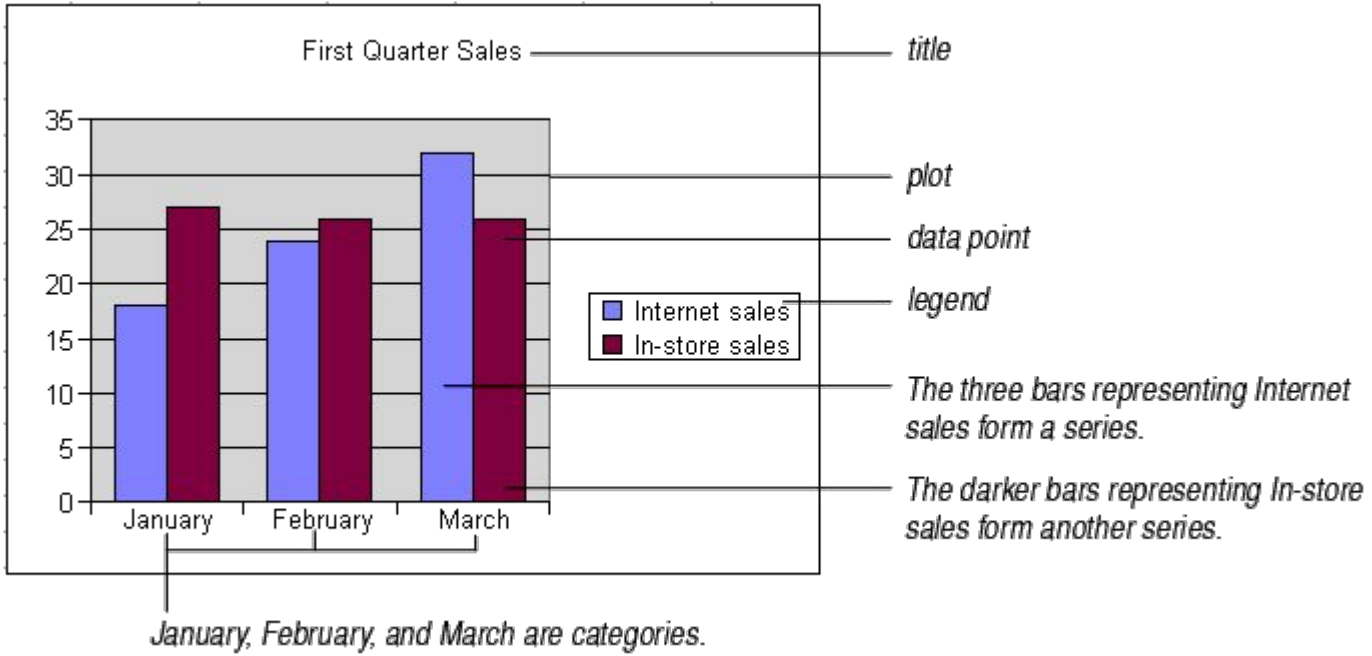
- Showing comparisons is really the basis of all good scientific investigation.
- Show causality, mechanism, explanation, systematic structure
- The real world is multivariate. For anything that you might study, there are usually many attributes that you can measure.
- Analytical presentations ultimately stand or fall depending on the quality, relevance, and integrity of their content.
- Integrate evidence
- Data graphics should be appropriately documented with labels, scales, and sources. A general rule for me is that a data graphic should tell a complete story all by itself.

Graphing Systems and Tools in Python

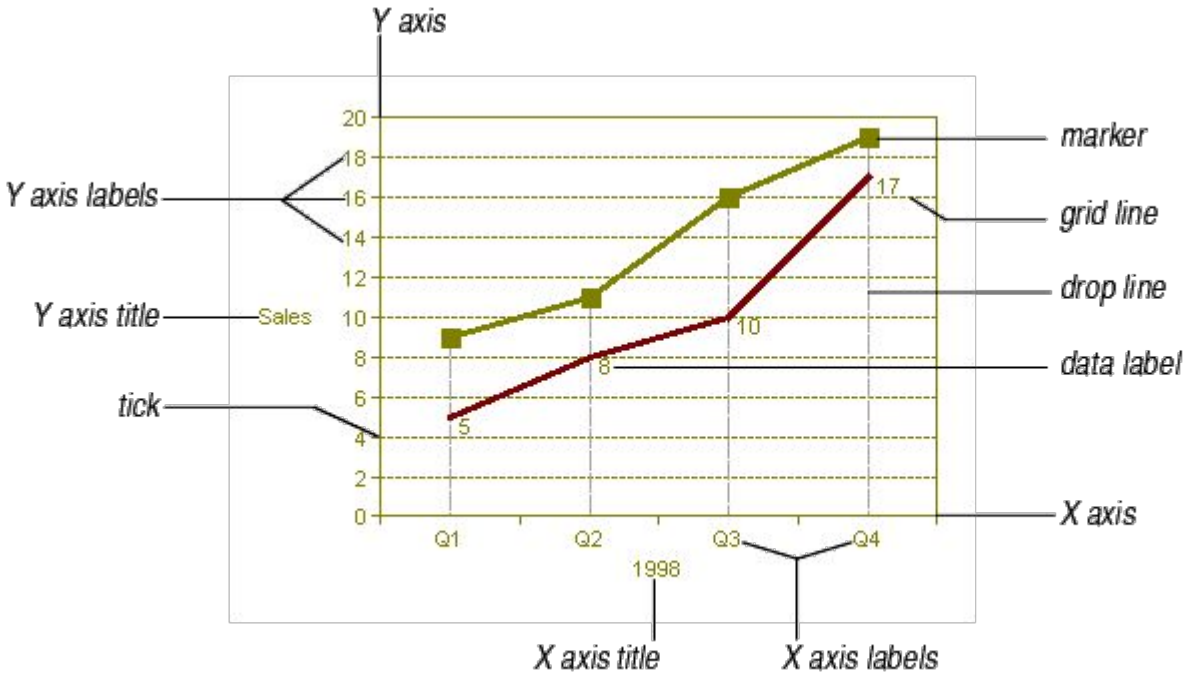
Basic Graphic Tools in Python

Characteristics	Matplotlib	Seaborn
Use Cases	Matplotlib plots various graphs using Pandas and Numpy	Seaborn is the extended version of Matplotlib which uses Matplotlib along with Numpy and Pandas for plotting graphs
Complexity of Syntax	It uses comparatively complex and lengthy syntax.	It uses comparatively simple syntax which is easier to learn and understand.
Multiple figures	Matplotlib has multiple figures can be opened	Seaborn automates the creation of multiple figures which sometimes leads to out of memory issues
Flexibility	Matplotlib is highly customizable and powerful.	Seaborn avoids a ton of boilerplate by providing default themes which are commonly used.

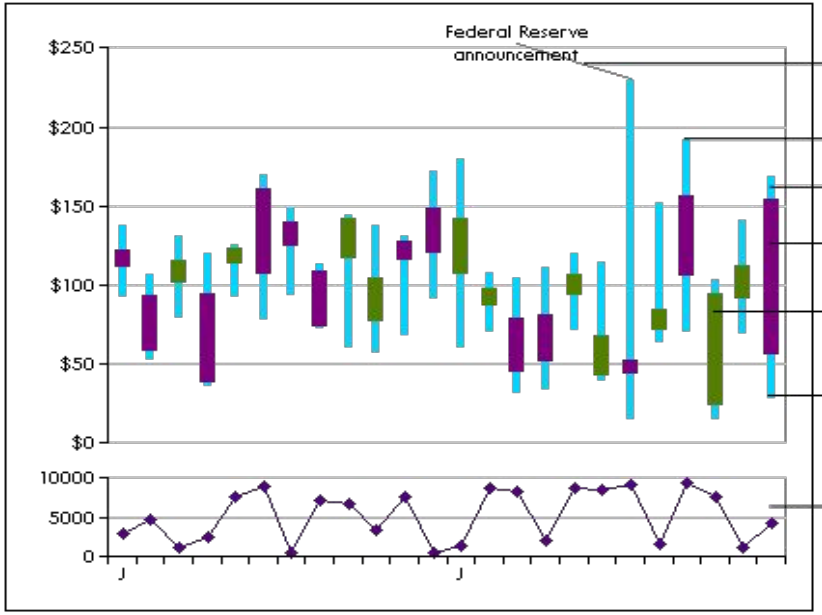
Elements of a Graph



Elements of a Graph



Elements of a Graph



label line

highest price attained for that period

high-low line

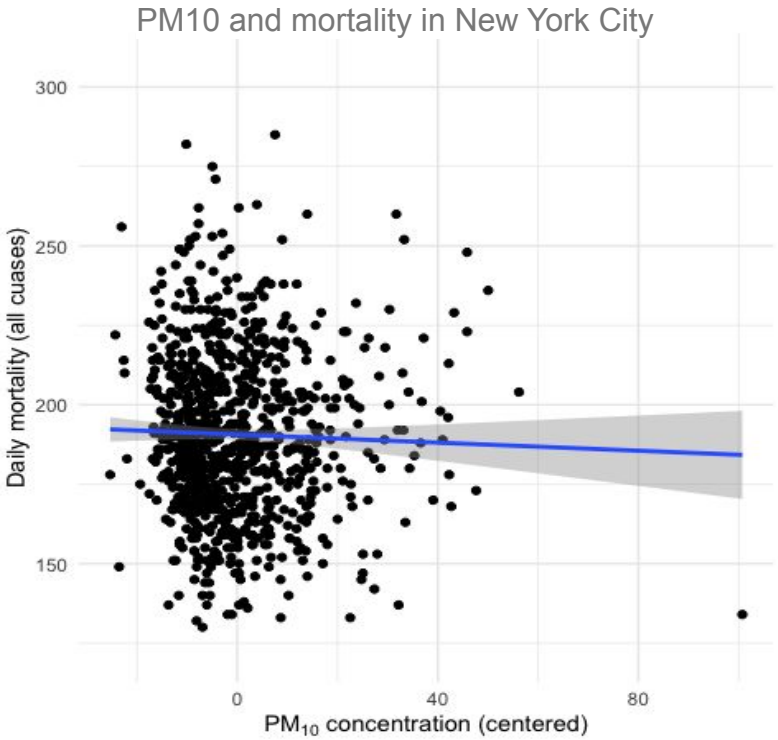
open-close bar showing gain

open-close bar showing loss

lowest price attained for that period

study showing volume data

Storytelling with Data - An Example



- **The Data** - Each point on the plot represents the average PM10 level for that day (measured in micrograms per cubic meter) and the number of deaths on that day. The PM10 data come from the U.S. Environmental Protection Agency and the mortality data come from the U.S. National Center for Health Statistics.
- **The Analysis** - This is a bivariate plot showing two variables in this dataset. From the plot it seems that there is a slight negative relationship between the two variables.
- **The Insight** - That is, higher daily average levels of PM10 appear to be associated with lower levels of mortality (fewer deaths per day).

Conclusion

Benefits of Data Visualization

- Better monitoring
- Taking decisions
- Helps business intelligence
- Accurate data
- Finding efficiencies/ inefficiencies