11 most important plots in data science in a single frame.  
.  
.  
Visualizations are critical in understanding complex data patterns and relationships.  
  
They offer a concise way to understand the:  
- intricacies of statistical models  
- validate model assumptions  
- evaluate model performance, and much more.  
  
The visual below depicts the 11 most important and must-know plots in data science:  
  
👉 Find a more vivid explanation with visuals here: <https://bit.ly/ds-plots>  
  
1) KS Plot:  
- It is used to assess the distributional differences.  
- The core idea is to measure the maximum distance between the cumulative distribution functions (CDF) of two distributions.  
- The lower the maximum distance, the more likely they belong to the same distribution.  
- Thus, instead of a “plot”, it is mainly interpreted as a “statistical test” to determine distributional differences.  
  
2) SHAP Plot:  
- It summarizes feature importance to a model’s predictions by considering interactions/dependencies between them.  
- It is useful in determining how different values (low or high) of a feature affect the overall output.  
  
3) ROC Curve:  
- It depicts the tradeoff between the true positive rate (good performance) and the false positive rate (bad performance) across different classification thresholds.  
  
4) Precision-Recall Curve:  
- It depicts the tradeoff between Precision and Recall across different classification thresholds.  
  
5) QQ Plot:  
- It assesses the distributional similarity between observed data and theoretical distribution.  
- It plots the quantiles of the two distributions against each other.  
- Deviations from the straight line indicate a departure from the assumed distribution.  
  
6) Cumulative Explained Variance Plot:  
- It is useful in determining the number of dimensions we can reduce our data to while preserving max variance during PCA.  
  
7) Elbow Curve:  
- The plot helps identify the optimal number of clusters for the k-means algorithm.  
- The point of the elbow depicts the ideal number of clusters.  
  
8) Silhouette Curve:  
- The Elbow curve is often ineffective when you have plenty of clusters.  
- Silhouette Curve is a better alternative, as depicted above.  
  
9) Gini-Impurity and Entropy:  
- They are used to measure the impurity or disorder of a node or split in a decision tree.  
- The plot compares Gini impurity and Entropy across different splits.  
- This provides insights into the tradeoff between these measures.  
  
10) Bias-Variance Tradeoff:  
- It is used to find the right balance between the bias and the variance of a model against complexity.  
  
11) PDP:  
- Depicts the dependence between target and features.  
- A plot between the target and one feature forms → 1-way PDP.  
- A plot between the target and two feature forms → 2-way PDP.  
  
--  
👉 Get a Free Data Science PDF (550+ pages) by subscribing to my daily newsletter: <https://bit.ly/DailyDS>.  
--  
  
👉 Over to you: Which important plots have I missed here?  
  
[#machinelearning](https://www.linkedin.com/feed/hashtag/?keywords=machinelearning&highlightedUpdateUrns=urn%3Ali%3Aactivity%3A7131303006104391684)  
.

Activate to view larger image,

