# Data Studio Charting Checklist

You can see a few suggestions at <http://designingviz.com/> which may or may not be reasonable. I also really recommend The Wall Street Journal Guide to Information Graphics, a pretty tiny book that is remarkably helpful at helping you not screw things up.

Micro-tutorials for Illustrator can be found at <http://jonathansoma.com/lede/data-studio/>, including [how to open your Python files](http://jonathansoma.com/lede/data-studio/adobe-illustrator/opening-your-graph/) in Illustrator. Longer, detailed Illustrator tutorials can be found on Lynda.com, accessed for free through the [Columbia portal](https://ctl.columbia.edu/resources-and-technology/teaching-with-technology/tech-resources/lynda/).

## Communicating successfully

### Story

* What am I supposed to take away from the graphic?
* The graphic has a headline
* The graphic explains what the data is
* Typically, headline is editorialized and subhead explains what the data is
* There is a focus to the graphic, for example:
  + Something to obviously pay attention to
  + Something(s) in a highlight color
  + An annotated point on a line
* List the source of your data (usually in small text, bottom)
* Be honest mathematically: use per capita when necessary
* Can your graphic stand alone, outside of the story?

### Annotations

* What should someone be interested in? Mark interesting points, tell us a story.
* Don't label every point, no one will pay attention
* But if you have a lot of points or bars or something, you should probably label some of them so people know what to look at
* Highlight background areas to talk about things that happen over time
* Use a single line to talk about something that happened at one point in time
* Do you draw a line to the interesting point? Or an arc? Or just put the text next to it?

### Busy-ness

* Clear hierarchy of what to pay attention to (more important = bigger, more colorful)
* Do you have too many labels?
* Do you have too many bars/lines?
* You can combine smaller categories into an 'other' category if necessary
* You don't have to show all of your datapoints, subsets or top/bottom is usually fine
* No one knows what error bars, margin of error, or confidence intervals are. If you're using them, have a good reason
* Matplotlib likes to put boxes around all of your charts. You probably can open them up.

## Visual design

### Color

* Use nice colors! Steal a scheme from somewhere. Colorbrewer2.org is basic but good.
* Things that aren’t important become grey, things that are important become bright/dark
* Only using different colors if the colors mean something
* Are you using the right kind of color scale?
  + Categorical for categories (e.g. different types of crimes)
  + Sequential for ordered numbers or categories (e.g. higher heat is darker red)
  + Diverging for moving away from a middle ground (e.g. voting Republican/Democrat, more/less red/blue)
* Matching colors between text and elements on the page can look nice if they’re talking about the same thing
* Legends and keys
  + Instead of using a legend, can you directly label what the points/colors/etc are?
  + Have your color categories be round numbers, not "dark blue is 145.4-156.2"

### Axes

* If years, not every year should be marked
  + if pandas is doing it, try converting year to integer before plotting
* Ticks and grid lines at round numbers
* No sideways labels!
* No labels at all, if it can be avoided
* Units attached to first/line number on axis
* Do you really need to label that 0? Do you really need the highest label on the axis?
* If it's money, you'll need a currency sign somewhere
* Do you need that axis line? Maybe, maybe not.
  + Sometimes grid lines are better!
* Don't repeat yourself with labels
* Is the maximum and minimum reasonable?

### Text

* Pick nice fonts - search online for font combinations you like
* Only a few sizes of text (e.g. title, subhead, axes, annotations)
* Large enough and readable
* Clear hierarchy between text elements - title > subhead > annotations > axes
* Translate weird phrases or jargon from your dataset into "real people" words
* Don't put background colors on text elements
* Commas in thousands
* Don't repeat yourself with labels

### Specific Charts

* Bars/columns
  + Order them by size
  + Should not have grid lines going in same direction as bars
  + Use labels on the bars instead of an axis if there aren't very many
  + No little tick line at the beginning of the bar (matplotlib and Illustrator both love to do this)
* Line charts
  + If you're doing lines + dots on measurements, do you really need the dots?
  + Area charts for "stock" and line charts for "flow". Area means “stuff,” basically.
* Maps
  + Only if there's a geographic trend (e.g. east coast looks different than west coast)
  + If it's the USA: Albers projection, not Mercator
* Pie chart
  + Are you sure? Maybe do a single stacked bar instead, they look nicer (you can even do it in Illustrator real easy)
  + Four or fewer slices
  + Start with the biggest slice at twelve o’clock, then go in order
* Small multiples
  + Each multiple is the same size with same axes
  + Each multiple is actually small
  + People can understand the measurements and axes