Appendix A: Resources And Tools

While we’ve provided contextual links and resources throughout the book, there were a few that didn’t fit properly in the chapters but are “go to” resources that are either part of our daily workflows. We’ve compiled them—along with a “best of the best” of links from selected chapters—into an organized and annotated list for quick reference.

Data Cleansing

* *OpenRefine* (<http://openrefine.org/>) – an open source, locally-installed, cross-platform toolkit that makes it extremely easy to import, explore, clean, transform, and enrich messy data into something usable for analysis.
* *DataWrangler* (<http://vis.stanford.edu/wrangler/>) – An browser-based, JavaScript tool created by Stanford University’s Visualization Group that lets you explore and transform small data sets in-browser, then export a custom Python or JavaScript source file, suitable for running locally on both small and large data sets.
* *WebPlotDigitizer* (<http://arohatgi.info/WebPlotDigitizer/app/>) – This online tool makes it possible to quickly “reverse engineer” charts and graphs that have no associated open data files.
* *Google CRUSH Tools* (<https://code.google.com/p/crush-tools/>) – A command-line processing engine and data transformation tool that makes it possible to work efficiently with large data sets from a shell prompt.
* *csvkit* (<https://github.com/onyxfish/csvkit>) – A suite of open source Python utilities that are similar to the *CRUSH* tools, but usable from both the command line and within Python scripts.
* *Data Cleaner* (<http://datacleaner.org/>) – This product is similar to OpenRefine but with both commercial and open source offerings.
* *Mr. Data Converter* (<http://shancarter.github.io/mr-data-converter/>) – In-browser and locally installable open source tool created by Shan Carter to improve data cleansing workflows at the New York Times.
* *Miso Dataset* (<http://misoproject.com/dataset/>) – Client-side JavaScript data transformation and management library.
* *Your favorite scripting language* – Never underestimate the power of a Python, R, Perl or awk script when it comes to cleaning data. You’ll have to do more up-front work, but you may be able to build a far more reusable and customized cleanup and transformation workflow with your own tools.

Data Analytics & Visualization – Core Tools

* R (<http://www.r-project.org/>) + RStudio (<http://www.rstudio.com/>) – *The* language of data science. Commercial offering available via Revolution Analytics (<http://www.revolutionanalytics.com/>).
* Python (<http://www.python.org/>) + pandas (<http://pandas.pydata.org/>) – The *other* language of data science. Additional open source and commercial offerings available via *Enthought Canop*y (<https://www.enthought.com/products/canopy/>) and *Continuum Analytics Anaconda* (<http://docs.continuum.io/anaconda/install.html>).
* *Tableau* (<http://www.tableausoftware.com/>) – Commercial tool with an emphasis on producing interactive dashboards and visualizations.

Data Analytics & Visualization – JavaScript Tools

* *D3.js* (<http://d3js.org/>) – Enables the creation of “data driven documents” and provides templates and examples for creating almost every type of modern static and interactive visualization.
* *JavaScript InfoVis Toolkit* (<http://philogb.github.io/jit/>) – Similar to D3, but may be more accessible to those new to JavaScript
* *Highcharts JS* (<http://www.highcharts.com/>) – Provides robust charting and graphing funtions, especially well-suited for dashboards.

Data Analytics & Visualization – Mapping Tools

* *OpenHeatMap* (<http://www.openheatmap.com/>) – Produce high quality heat maps from CSV data right in your browser. No coding required.
* *Leaflet* (<http://leafletjs.com/>) – A very robust and mobile-friendly, JavaScript mapping library.

Data Analytics & Visualization – Specialized Tools

* *TimeFlow* (<https://github.com/FlowingMedia/TimeFlow/wiki>) – An open source tool specifically design for analysis and visualization of temporal/time series data.
* *Gephi* (<https://gephi.org/>) – Open source network graph analysis and visualizatin tool.
* *Quadirgram* (<http://www.quadrigram.com/>) – This tool provides a visual programming interface for working with data and designing highly customized, interactive visualization.

Aggregation Sites, Q&A Sites, And Blogs To Follow

* *R-Bloggers* (<http://www.r-bloggers.com/>) – Rather than follow a plethora of individual blogs you can follow the R-Bloggers RSS feed to see only R-related posts that deal with all aspects of data analysis and visualization.
* *Stats Blogs* (<http://www.statsblogs.com/>) – An aggregation sites, similar to R-Bloggers, but with a focus on statistics.
* *StackExchange* (<http://stackexchange.com/>) – The perfect place to go if you have R, Python or pandas questions, can’t remember a ggplot option or need some help with a gnarly statistics problem.
* *JunkCharts* (<http://junkcharts.typepad.com/>) – Learn from the visualization mistakes of others.
* *FlowingData* (<http://flowingdata.com/>) – Resources, news and tutorials that will improve the way you think and design visualizations.
* *DataVisualization.ch* (<http://selection.datavisualization.ch/>) – Aggregation and index of the most popular and useful visualization tools currently available.
* *Data Analysis & Visualization Bit.ly Bundle* (<http://bitly.com/bundles/hrbrmstr/1>) – An aggregation of links maintained by the authors along with David Severski.

Color

* *ColorBrewer* (<http://colorbrewer2.org/>) – Designed by Cynthia brewer, this is *the* color resource that should be the first tool you head for when designing visualizations. It provides a wide range of palettes with options for creating print-safe and colorblind-friendly images.
* *HCL Picker* (http://tristen.ca/hcl-picker/) – An open source, D3-based color picker, that lets you select colors based on hue, chroma and lightness.
* *Adobe Kuler* (<https://kuler.adobe.com/>) – An online tool, provided by Adobe, which allows you to design compelling color palettes or choose from a wide assortment of pre-made palettes.
* *OS X Color Picker Palettes* (<https://github.com/sathomas/colors>) – Use *ColorBrewer* palettes in Excel, PhotoShop and any other application on your Mac.