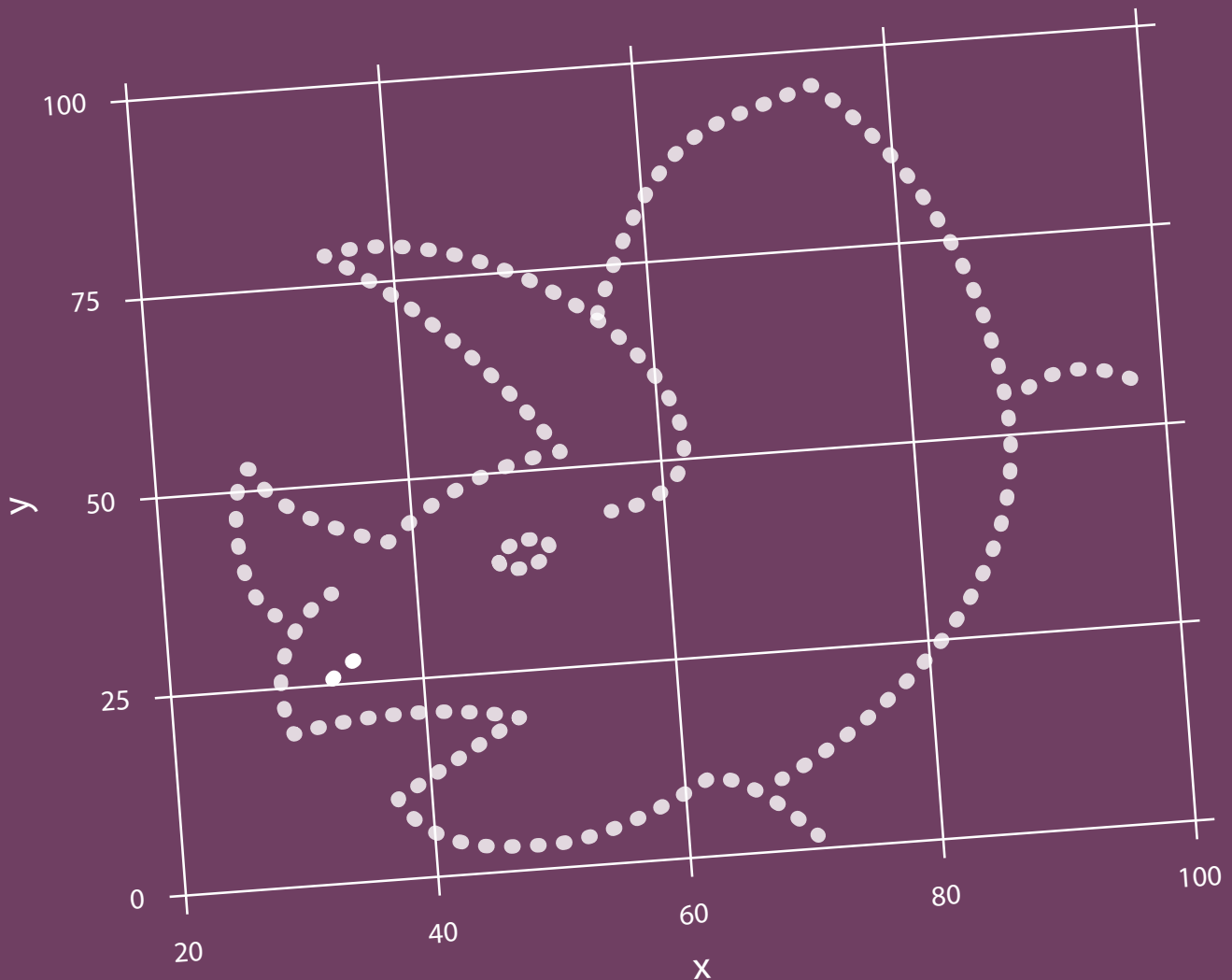


THE HITCHHIKER'S GUIDE TO PLOTNINE

Don't panic and create beautiful plots with Python



Jodie Burchell
Mauricio Vargas

The Hitchhiker's Guide to Plotnine

Don't panic and create beautiful graphs with Python

by

Jodie Burchell, Mauricio Vargas

The Hitchhiker's Guide to Plotnine

Jodie Burchell & Mauricio Vargas

This book is for sale at <https://leanpub.com/plotnine-guide>

This version was published on 28th October, 2019.

©2019 Jodie Burchell & Mauricio Vargas



This is a Leanpub book. Leanpub empowers authors and publishers with the Lean Publishing process. Lean Publishing is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

Contents

What to expect from this book	1
1 Line plots	3
1.1 Introduction	3
1.2 Basic ggplot structure	4
1.3 Basic line plot	5
1.4 Adjusting line width	6
1.5 Changing group labels	7
1.6 Adjusting the axis scales	8
1.7 Adjusting axis labels & adding title	9
1.8 Adjusting the colour palette	10
1.9 Adjusting the legend	12
1.10 Using the white theme	16
1.11 Creating an XKCD style chart	17
1.12 Using the 'Five Thirty Eight' theme	19
1.13 Creating your own theme	21
2 Area plots	25
2.1 Introduction	25
2.2 Basic ggplot structure	26
2.3 Basic area plot	27
2.4 Changing group labels	29
2.5 Adjusting the axis scales	30
2.6 Adjusting axis labels & adding title	31
2.7 Adjusting the colour palette	31
2.8 Adjusting the legend	34
2.9 Using the white theme	37
2.10 Creating an XKCD style chart	38
2.11 Using the 'Five Thirty Eight' theme	41
2.12 Creating your own theme	42
3 Bar plots	45
3.1 Introduction	45
3.2 Basic ggplot structure	46

Contents

3.3	Basic bar plot	47
3.4	Adding data labels	49
3.5	Adjusting data labels position	49
3.6	Changing group labels	50
3.7	Adjusting the axis scales	51
3.8	Adjusting axis labels & adding title	52
3.9	Adjusting the colour palette	53
3.10	Adjusting the legend	55
3.11	Using the white theme	59
3.12	Creating an XKCD style chart	60
3.13	Using the 'Five Thirty Eight' theme	63
3.14	Creating your own theme	64
4	Stacked bar plots	67
4.1	Introduction	67
4.2	Basic ggplot structure	68
4.3	Basic stacked bar plot	69
4.4	Adding data labels	71
4.5	Adjusting data labels position	71
4.6	Changing group labels	72
4.7	Adjusting the axis scales	73
4.8	Adjusting axis labels & adding title	74
4.9	Adjusting the colour palette	75
4.10	Adjusting the legend	78
4.11	Using the white theme	81
4.12	Creating an XKCD style chart	82
4.13	Using the 'Five Thirty Eight' theme	85
4.14	Creating your own theme	87
5	Scatterplots	90
5.1	Introduction	90
5.2	Basic ggplot structure	92
5.3	Basic scatterplot	93
5.4	Changing the shape of the data points	94
5.5	Adjusting the axis scales	95
5.6	Adjusting axis labels & adding title	96
5.7	Adjusting the colour palette	97
5.8	Adjusting the legend	104
5.9	Using the white theme	107
5.10	Creating an XKCD style chart	108
5.11	Using the 'Five Thirty Eight' theme	111
5.12	Creating your own theme	112
6	Weighted scatterplots	115

Contents

6.1	Introduction	115
6.2	Basic ggplot structure	117
6.3	Basic weighted scatterplot	118
6.4	Changing the shape of the data points	119
6.5	Adjusting the axis scales	122
6.6	Adjusting axis labels & adding title	123
6.7	Adjusting the colour palette	123
6.8	Adjusting the size of the data points	130
6.9	Adjusting the legend	132
6.10	Using the white theme	135
6.11	Creating an XKCD style chart	136
6.12	Using the 'Five Thirty Eight' theme	139
6.13	Creating your own theme	141
7	Histograms	144
7.1	Introduction	144
7.2	Basic ggplot structure	145
7.3	Basic histogram	146
7.4	Adding a normal density curve	147
7.5	Changing from density to frequency	148
7.6	Adjusting binwidth	149
7.7	Adjusting the axis scales	150
7.8	Adjusting axis labels & adding title	151
7.9	Adjusting the colour palette	152
7.10	Using the white theme	156
7.11	Creating an XKCD style chart	157
7.12	Using the 'Five Thirty Eight' theme	159
7.13	Creating your own theme	160
7.14	Adding lines	162
7.15	Multiple histograms	163
7.16	Formatting the legend	166
8	Density plots	168
8.1	Introduction	168
8.2	Basic ggplot structure	169
8.3	Basic density plot	170
8.4	Adjusting the axis scales	171
8.5	Adjusting axis labels & adding title	172
8.6	Adjusting the colour palette	172
8.7	Using the white theme	175
8.8	Creating an XKCD style chart	176
8.9	Using the 'Five Thirty Eight' theme	178
8.10	Creating your own theme	180
8.11	Adding lines	181

Contents

8.12	Multiple density plots	182
8.13	Formatting the legend	187
9	Function plots	189
9.1	Introduction	189
9.2	Basic ggplot structure	190
9.3	Plotting a normal curve	191
9.4	Plotting a t-curve	192
9.5	Plotting your own function	192
9.6	Plotting multiple functions on the same graph	193
9.7	Adjusting the axis scales	194
9.8	Adjusting axis labels & adding title	195
9.9	Changing the colour of the function lines	196
9.10	Adding a legend	198
9.11	Changing the size of the lines	200
9.12	Using the white theme	201
9.13	Creating an XKCD style chart	202
9.14	Using the 'Five Thirty Eight' theme	205
9.15	Creating your own theme	207
9.16	Adding areas under the curve	209
10	Boxplots	212
10.1	Introduction	212
10.2	Basic ggplot structure	213
10.3	Basic boxplot	214
10.4	Customising axis labels	215
10.5	Changing axis ticks	216
10.6	Adding a title	217
10.7	Changing the colour of the boxes	218
10.8	Using the white theme	222
10.9	Creating an XKCD style chart	223
10.10	Using the 'Five Thirty Eight' theme	225
10.11	Creating your own theme	227
10.12	Boxplot extras	228
10.13	Grouping by another variable	230
10.14	Formatting the legend	233
11	Linear regression plots	235
11.1	Introduction	235
11.2	Basic ggplot structure	237
11.3	Trend line plot	238
11.4	Customising axis labels	241
11.5	Adding a title	241
11.6	Adjusting the axis scales	242

Contents

11.7	Using the white theme	244
11.8	Creating an XKCD style chart	245
11.9	Using the 'Five Thirty Eight' theme	247
11.10	Creating your own theme	249
11.11	Regression diagnostics plots	251
11.11.1	Residual vs fitted plot	252
11.11.2	Normal Q-Q plot	254
11.11.3	Scale-Location plot	255
11.11.4	Residuals vs Leverage plot	257
11.12	Customising the residual plots	258
11.13	Combining the plots into a single plot	264
12	LOWESS plots	267
12.1	Introduction	267
12.2	Basic ggplot structure	268
12.3	Creating a basic LOWESS plot, and what it can tell us about our data	269
12.4	Changing the width of the bins	271
12.5	Adjusting the axis scales	272
12.6	Adjusting axis labels & adding title	273
12.7	Changing the colour and size of the LOWESS curve	274
12.8	Changing the appearance of the confidence interval	277
12.9	Changing the appearance of the scatterplot	280
12.10	Using the white theme	283
12.11	Creating an XKCD style chart	284
12.12	Using the 'Five Thirty Eight' theme	286
12.13	Creating your own theme	288
Further reading		290
Plotnine		290
Python 3		290
Pandas		290

What to expect from this book

As users of R, we are both massive fans of the elegant `ggplot2` which makes it simple to create beautiful plots with just a few lines of code. However, when moving to Python we found that `matplotlib` was not quite as intuitive or easy to use. Fortunately, `plotnine` has been created¹, which is a stable and comprehensive port of `ggplot2` to Python. `plotnine` sits over `matplotlib`, which allows it to use the graphing functionality already built into Python with the grammar of `ggplot` to give it a much greater ease of use.

This book was inspired by our previous book² on using `ggplot` in R; however, the code in this book is completely new and tailored for using `plotnine` in Python. The focus of the book is technical, and aims to get you to create your own customised graphs in `plotnine` as quickly as possible. In this sense, the book resembles a recipe book with detailed instructions on how to make each change and customisation in your own plots. It is assumed that you know the basics of Python and want to focus on making beautiful plots with `plotnine`; as such, the basics of using Python or core packages such as `pandas` and `numpy` will not be covered in this book. However, we have included some resources in our further reading list if you need help getting started with Python and `pandas` more generally.

Each chapter will explain how to create a different type of plot, and will take you step-by-step from a basic plot to a highly customised graph. Each chapter is independent from the others. You can read the whole book or go to a section of your interest, and we are sure that it will be easy to understand the instructions and reproduce our examples without reading the earlier chapters. Chapters 1-4 are based on an international trade dataset created from different sources (Chile Customs, Central Bank of Chile and General Directorate of International Economic Relations). The rest of the chapters are based in classic example datasets (`diamonds`, `galton` and `airquality`).

Digital formats allow us to update this book on a constant frequency. In this version we are using Python 3.7.1. To make sure you are using the same version

¹<https://plotnine.readthedocs.io/en/stable/>

²https://leanpub.com/hitchhikers_ggplot2

Contents

of the packages as this version of the book, the materials for this book also include a requirements file (`plotnine_requirements.txt`).

We invite you to stay in touch and read the authors' blogs where they publish articles about Python, R and Statistics. Jodie's blog is *Standard error*³ and Mauricio's blog is *Pachá (Batteries Included)*⁴. We'd also like to thank the talented Aga Kozmic for designing the cover of this book. You can see more of her work on her website⁵.

Don't panic!

³<http://t-redactyl.io/>

⁴<https://pacha.hk/>

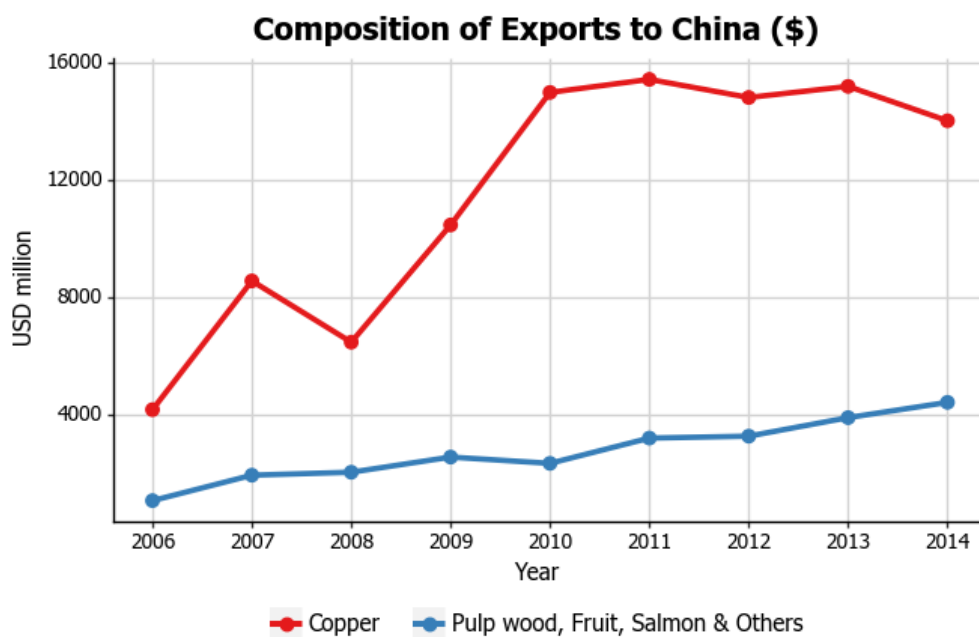
⁵<https://agakozmic.com/>

Chapter 1

Line plots

1.1 Introduction

In this chapter, we will work towards creating the line plot below. We will take you from a basic line plot and explain all the customisations we add to the code step-by-step.



Chapter 1 Line plots

The first step is to import all of the required packages. For this we need:

- pandas and its DataFrame class to read in and manipulate our data;
- plotnine to get our data and create our graphs; and
- numpy to do some basic numeric calculations in our graphing.

We can also change the size of the plots using the `figure_size` function from plotnine. We have resized the plots in this chapter so they display a little more neatly.

```
import numpy as np
import pandas as pd

import plotnine
plotnine.options.figure_size=(7.5, 4.2)

from plotnine import *
from pandas import DataFrame
```

For this chapter, we'll be using a trade dataset put together by the book's authors. The exact sample we've used in this chapter can be downloaded from [here](https://git.io/JecIS)¹. You can load the data into Python like so:

```
copper = pd.read_csv("https://git.io/JecIS")
```

1.2 Basic ggplot structure

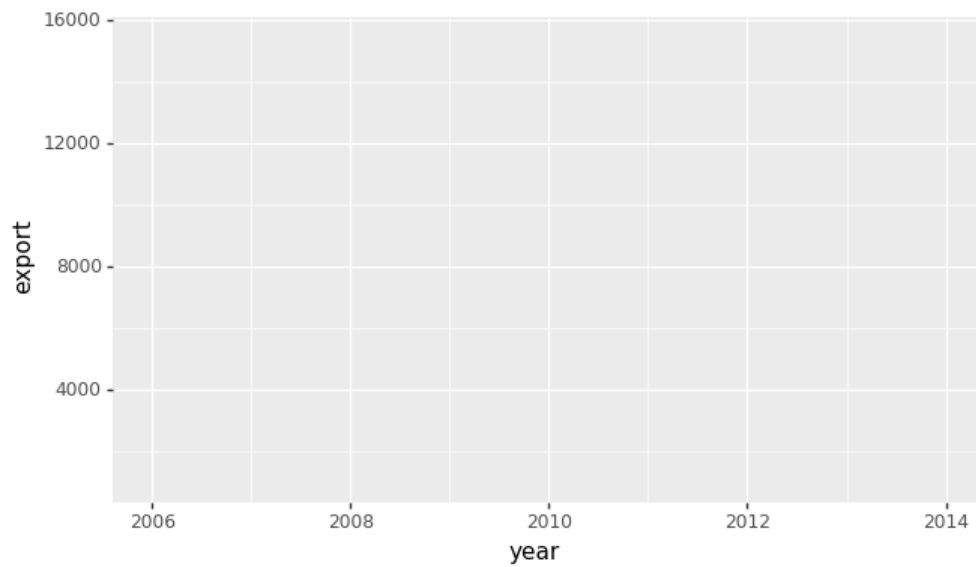
In order to initialise a line plot we tell ggplot that `copper` is our data, and specify that our x-axis plots the `year` variable, our y-axis plots the `export` variable, and our grouping variable is called `product`. You may have noticed that we put our variables inside a method called `aes`. This is short for aesthetic mappings, and determines how the different variables you want to use will be mapped to parts of the graph. As you can see below, ggplot has mapped `year` to the x-axis and `export` to the y-axis. We'll see how it handles `product` in a minute.

You might have also noticed that there is nothing in the plot. In order to render our data, we need to tell ggplot how we want to visually represent it.

```
p1 = ggplot(copper, aes("year", "export", colour="product"))
p1
```

¹<https://git.io/JecIS>

Chapter 1 Line plots

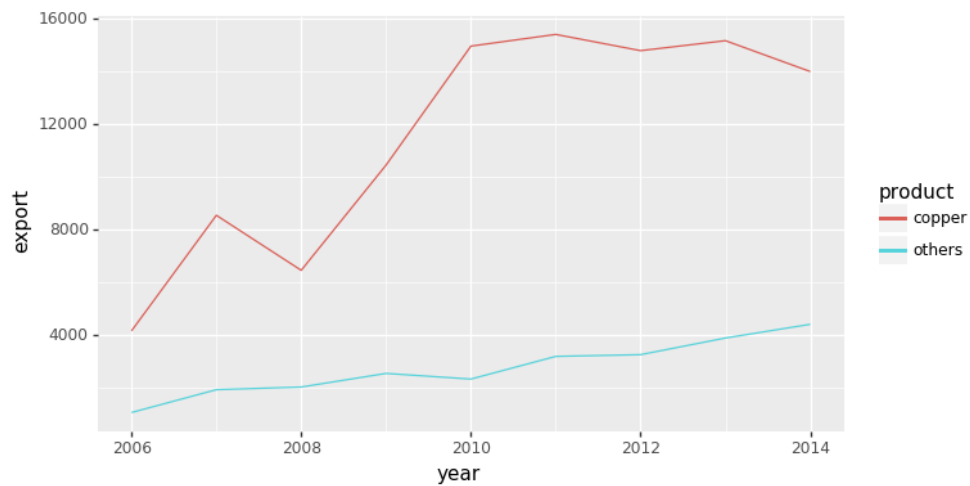


1.3 Basic line plot

We can do this using `geoms`. In the case of a line plot, we use the `geom_line()` geom.

```
p1 = (  
  ggplot(copper, aes("year", "export", colour="product"))  
    + geom_line()  
)  
p1
```

Chapter 1 Line plots



1.4 Adjusting line width

To change the line width, we add a size argument to `geom_line`.

```
p1 = (  
  ggplot(copper, aes("year", "export", colour="product"))  
  + geom_line(size=1.5)  
)  
p1
```

