Errata

Errata for *Python Data Science Handbook, Second Edition*

Print 🖨

<u>Submit your own errata for this product</u>.

| Confirmed Errata | Unconfirmed Errata |

The errata list is a list of errors and their corrections that were found after the product was released.

The following errata were submitted by our customers and have not yet been approved or disproved by the author or editor. They solely represent the opinion of the customer.

Color Key: Serious technical mistake   Minor technical mistake   Language or formatting error   Typo   Question   Note   Update

Version	Location	Description	Submitted by	Date submitted

Other Digital Version	Chap17 Loc7181 Partial slicing...	Kindle version: In[24]:poploc['california':'new york'] should be In[24]:pop.loc['California':'New York']	David Knuth	Feb 24, 2023
PDF	Page Exploring Modules with Tab Completion, Pg 9 6th para	Omit 'can be' in 'search can be useful'.	Harsh Shah	Apr 26, 2023
PDF	Page Pg 16, Underscore Shortcuts and Previous Outputs Last line	print(_) should print 1.0 instead of .0	Harsh Shah	Apr 27, 2023
Printed	Page page 180 Chapter 21 footer	Sex is not assigned at birth, it is defined at the moment of fertilization, so at birth it is only observed.	Michael Hare	Feb 12, 2024
ePub	Page Kindle Chap 20 pg 268 of 914 first code example - import seaborn as sns	The entire chapter is based on using the seaborn dataset on exoplanets. The code given to import this dataset does not work: planets = sns.load_dataset('planets'). Searching the web, there is a download at: github.com/mwaskom/seaborn-data/blob/master/planets.csv. Same issue with all of the downloads such as the US	Albert van Roodselaar	Jun 04, 2024

		You should have a link to errata and updated code for the epub versions of the book. Regards, Albert van Roodselaar		

Printed	Page page 189
	Figure 4-10
		Figure 4-10 — LOWESS target in partial-residual plot
		
		In the associated code in the GitHub repository for Figure 4-10, the LOWESS smoother is applied to the component (`results.ypartial`) rather than to the **partial residuals**. Replace:
		
		``python
		smoothed = sm.nonparametric.lowess(results.ypartial, results.feature, frac=1/3)
		``
		
		with:
		
		``python
		smoothed = sm.nonparametric.lowess(
		results.ypartial + results.residual, # PR_i(x)
		results.feature,
		frac=1/3
		)
		``
		
		Rationale. The gray dashed line is intended to be a LOWESS smooth of the **partial residuals** to show the empirical relationship without imposing a polynomial.
			Nahid Ahmadvand	Sep 14, 2025

PDF	Page 22 First paragraph second line	It is written as; "In addition to the enhanced interactive tools discussed in the previous chapter, Jupyter provides a number of ways to explore..." I think we have been discussing Ipython till here and it writes as; "Code development and data analysis always require a bit of trial and error, and IPython contains tools to streamline this process." in the following line. Thus, I think in the first sentence it should be <Ipython> instead of <Jupyter>.	Anonymous	Sep 30, 2025
PDF	Page 38 Caption of Figure 4-2	The caption of Figure 4-2 should be The difference between NumPy array and Python lists	ZHANG Hongyuan	Feb 17, 2023
PDF	Page 68 3rd paragraph (ignoring the codes)	"We see by rule 1 that the array a has fewer dimensions, so we pad it on the left with ones:" SHOULD BE: "We see by rule 1 that the array a has fewer dimensions, so we pad its shape on the left with ones:"	ZHANG Hongyuan	Feb 18, 2023
PDF	Page 84 Modifying Values with fancy indexing Second code snippet of the section	Out[19] should be [0 80 80 3 80 5 6 7 80 9]	Harsh Shah	Apr 30, 2023
PDF	Page 102; Boolean Operators Operator-Equivalend ufunc table	Missing symbol | for np.bitwise_or	Harsh Shah	Apr 30, 2023

		were "NaN" in Out[13] (which is incorrectly labeled as "Out[12]"). The call to mean() therefore doesn't appear to "fill with the mean of all values in A".		
Printed	Page 143 [38]	(I get the same result in Out[21] & Out[40] as in the text, showing that the pop Series is identical.) Results I get differ from the text as follows: In [38]: pop.unstack(level=0) Out[38]: state CA NY TX year 2010 37253956 19378102 25145561 2020 39538223 20201249 29145505	Gregory Sherman	Dec 21, 2023
Printed	Page 143 [39]	Result I get again differs from text: In [39]: pop.unstack(level=1) Out[39]: year 2010 2020 state California 37253956 39538223 New York 19378102 20201249 Texas 25145561 29145505	Gregory Sherman	Dec 21, 2023

Printed	Page 146 1st paragraph, 4th line of code	on page 146 there are differences in the code in the book and the code on the website: code in the book: ... <p style= 'font-family:"Courier New", Courier, monospace'>{0}{1} ... code on website: ...<p style='font-family:"Courier New", Courier, monospace'>{0}</p>{1} </div> ...	David Walden	Jan 17, 2024
Printed	Page 162 In[28]	I was not able to run this, but the form does not look right - it's df[a][b].unique(): In [28]: final['state'][final['area (sq. mi)'].isnull()].unique()	Gregory Sherman	Dec 23, 2023
Printed	Page 173 second sentence	"Here, because group A does not have a standard deviation ..." This does not match what is actually done: In[22]: def filter_func(x): return x['data2'].std() > 4 " The text should be something like "... column 'data2' of group A ...	Gregory Sherman	Dec 26, 2023

	Figure 26-3. A simple sinusoid via the object-oriented interface	interface. It uses the MATLAB interface, as shown in the code immediately above.		
PDF	Page 303 The paragraph one before the last.	"Minor ticks, though, have their labels formatted by a NullFormatter" The immediately preceding output example shows that LogFormatterSciNotation is used instead of NullFormatter.	Anonymous	Aug 13, 2023
PDF	Page 399 The last paragraph	This paragraph is the same as the last paragraph of p.396 and is misplaced in p.399.	Zhenhua Xu	Oct 26, 2023
PDF	Page 404 The last paragraph	This paragraph is the same as the last paragraph of p.396 and is misplaced in p.404.	Zhenhua Xu	Oct 26, 2023
Other Digital Version	518 1st paragraph	The "sex" column of the tips dataset is said to refer to the sex of the server: "... far more data on male servers ...". The R tips dataset, from which this dataset was sourced, attributes the "sex" column to the sex of the bill payer. (Kindle edition, p 518, location 15128)	John Winchester	Apr 18, 2023
Other Digital Version	526 Chapter 36, page 526 lower half	In the Kindle version in the chapter 36 example with marathon data, this line: data['split_sec'] = data['split'].view(int) / 1E9 should be: data['split_sec'] = data['split'].view('int64') / 1E9	David Knuth	Mar 09, 2023

Follow us   in   ▶

About O'Reilly

Teach/Write/Train

Careers

O'Reilly News

Media Coverage

Community Partners

Affiliate Program

Submit an RFP

Diversity

Content Sponsorship

Support

Contact Us

Newsletters

Privacy Policy

AI Policy

International

Australia & New Zealand

Japan

Download the O'Reilly App

Take O'Reilly with you and learn anywhere, anytime on your phone and tablet.

[Download on the App Store]   [GET IT ON Google Play]

Watch on Your Big Screen

View all O'Reilly videos, virtual conferences, and live events on your home TV.

[Roku Players & TVs]   [available at amazon appstore]

O'REILLY®

O'REILLY®

Explore Skills ⌄   Features ⌄   Radar Blog   Buy Courses

Plans   Sign In   Try Now