REPOSITORY LINE LENGTH ANALYSIS INSIGHTS INTO PYTHON PACKAGES

ROSEN
empowered by technology

ABOUT ME

Peer Wagner

Computer Science

@wagnerpeer

pwagner@rosen-group.com

Software Engineer

https://github.com/wagnerpeer

Slides: https://github.com/rosen-group/conferences

This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

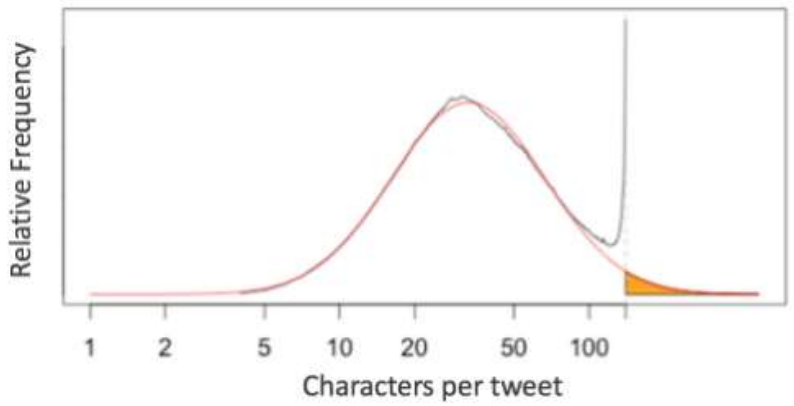
CHARACTER LIMIT TWITTER

ROSEN
empowered by technology

Exploring Line Lengths in Python Packages

Thu 09 November 2017

This week, Twitter upped their single-tweet character limit from 140 to 280, purportedly based on this interesting analysis of tweet lengths published on Twitter's engineering blog. The gist of the analysis is this: English language tweets display a roughly log-normal distribution of character counts, except near the 140-character limit, at which the distribution spikes:



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

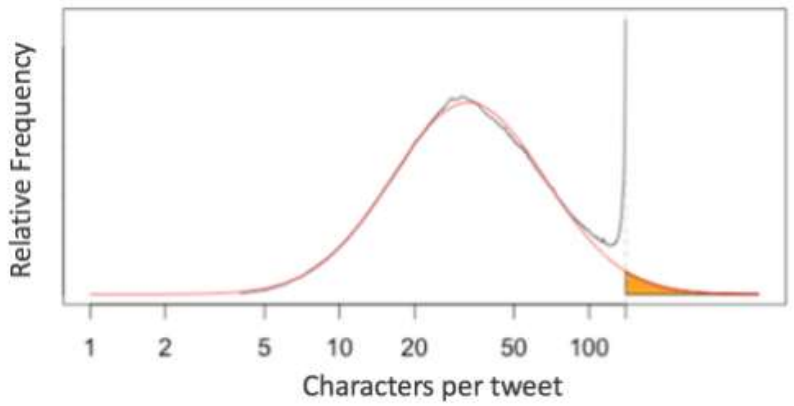
CHARACTER LIMIT
~~TWITTER~~ PYTHON PACKAGES

ROSEN
empowered by technology

Exploring Line Lengths in Python Packages

Thu 09 November 2017

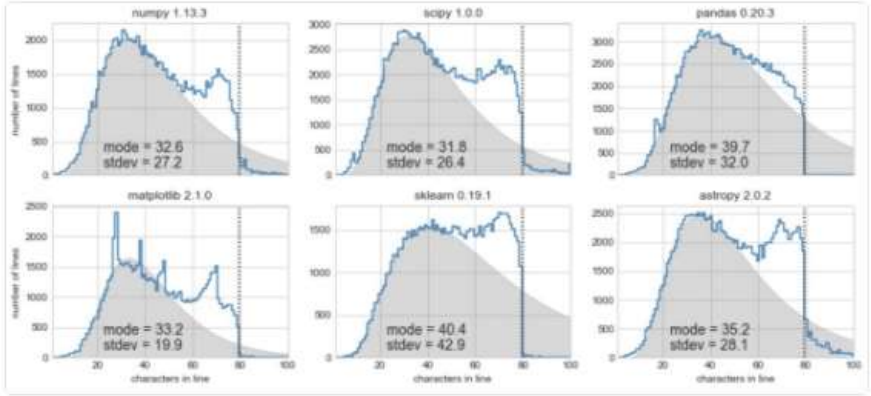
This week, Twitter upped their single-tweet character limit from 140 to 280, purportedly based on this interesting analysis of tweet lengths published on Twitter's engineering blog. The gist of the analysis is this: English language tweets display a roughly log-normal distribution of character counts, except near the 140-character limit, at which the distribution spikes:



Jake VanderPlas
@jakevdp

New post: some data-driven ruminations on PEP8's 79 character limit & the mark it leaves on Python packages
jakevdp.github.io/blog/2017/11/0 ...

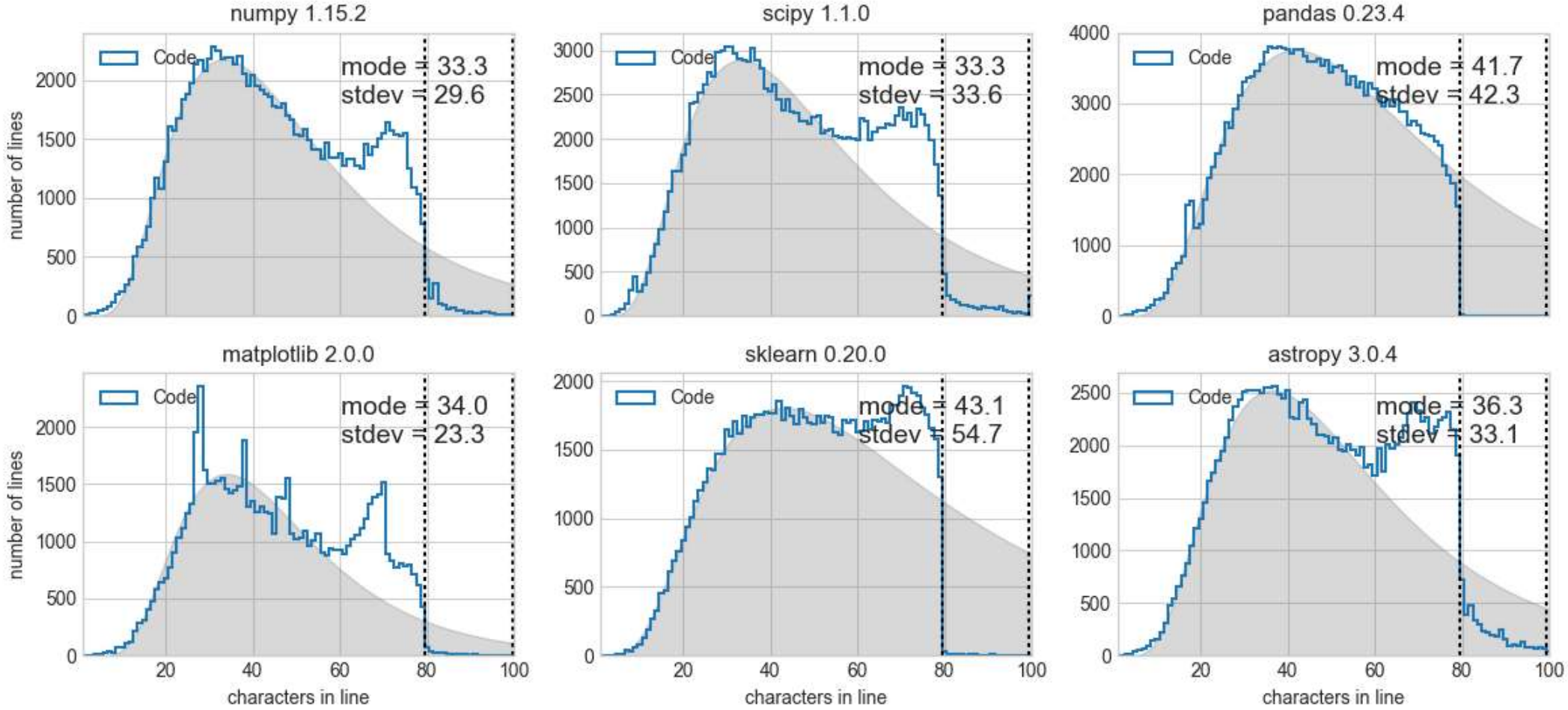
Tweet übersetzen



15:02 - 9. Nov. 2017

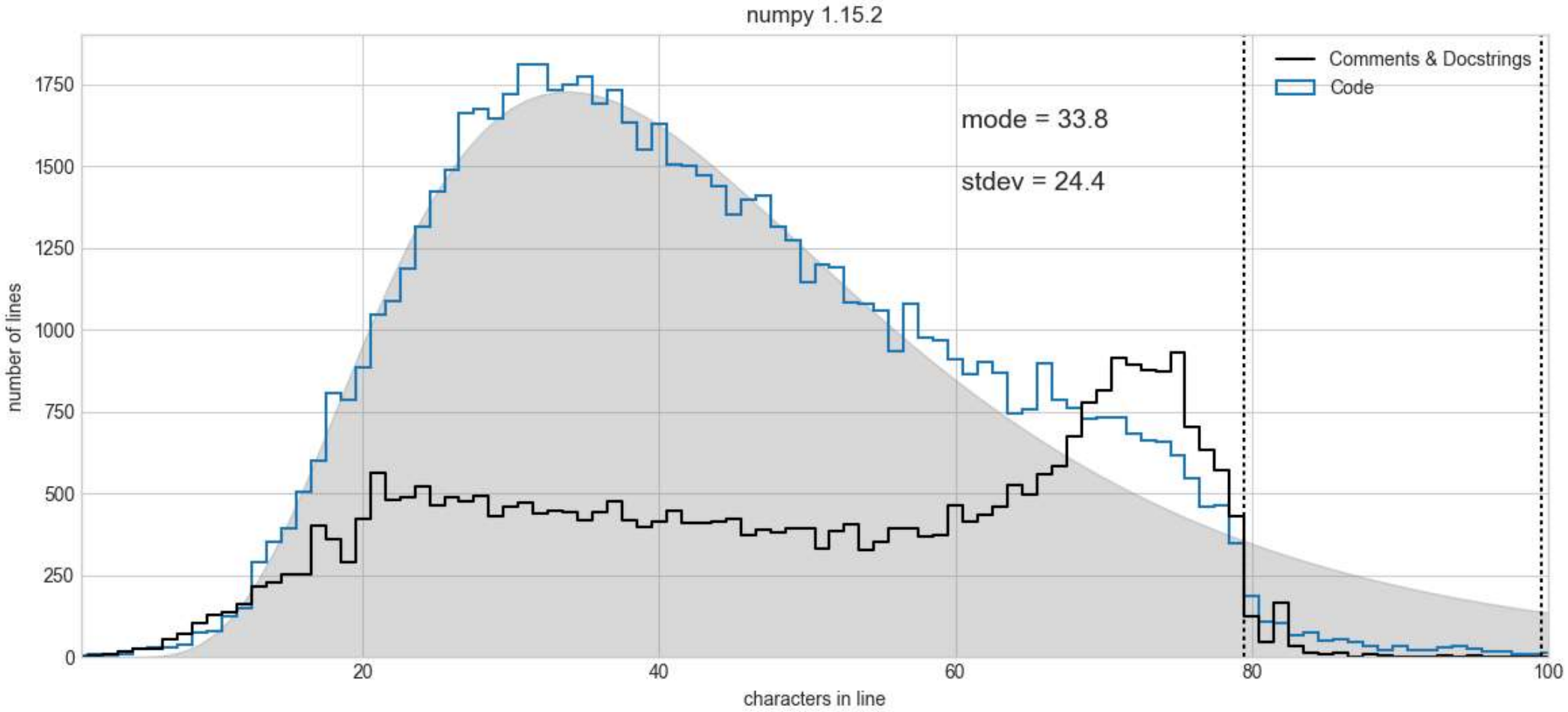
This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

REPRODUCTION OF JAKES RESULTS OVERVIEW



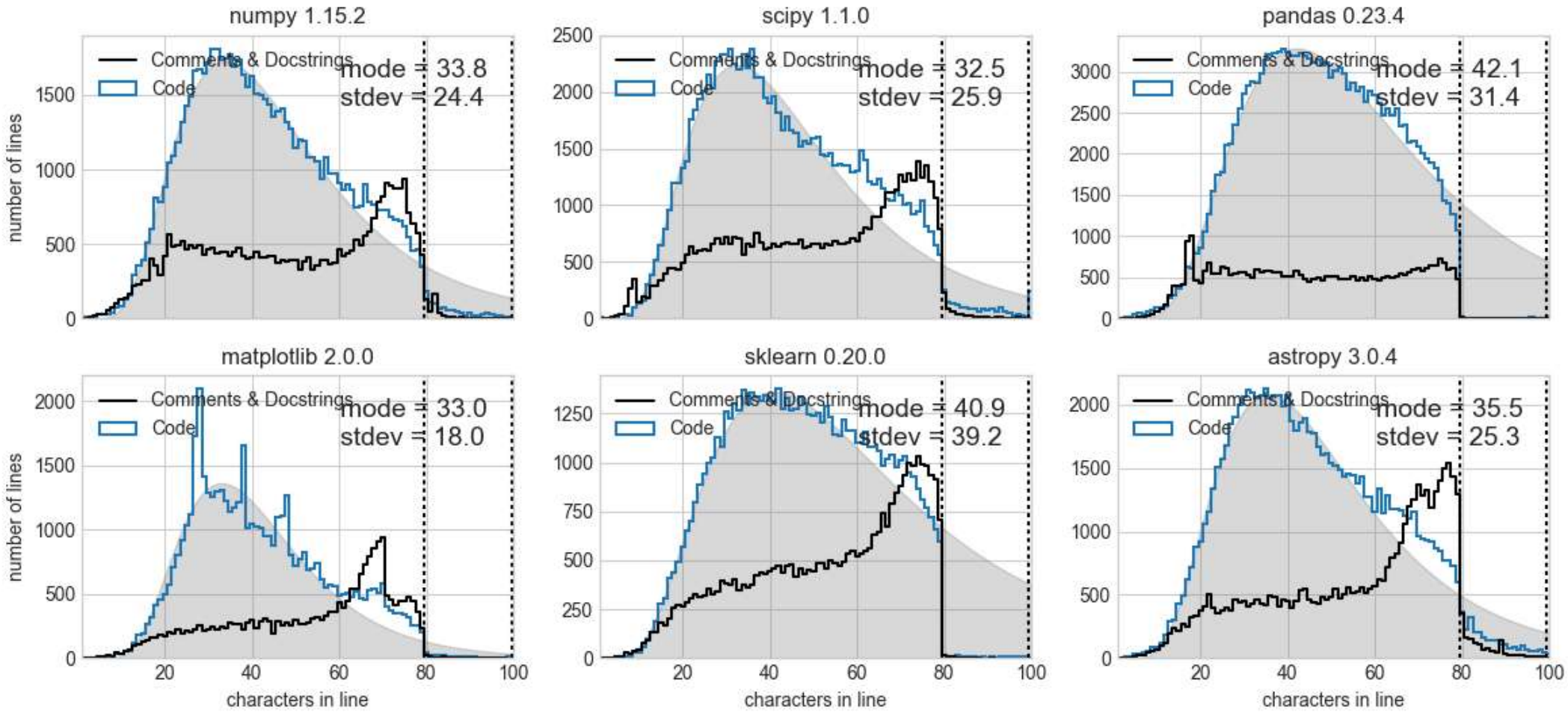
This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

COMMENTS & DOCSTRINGS
WHAT CHANGES



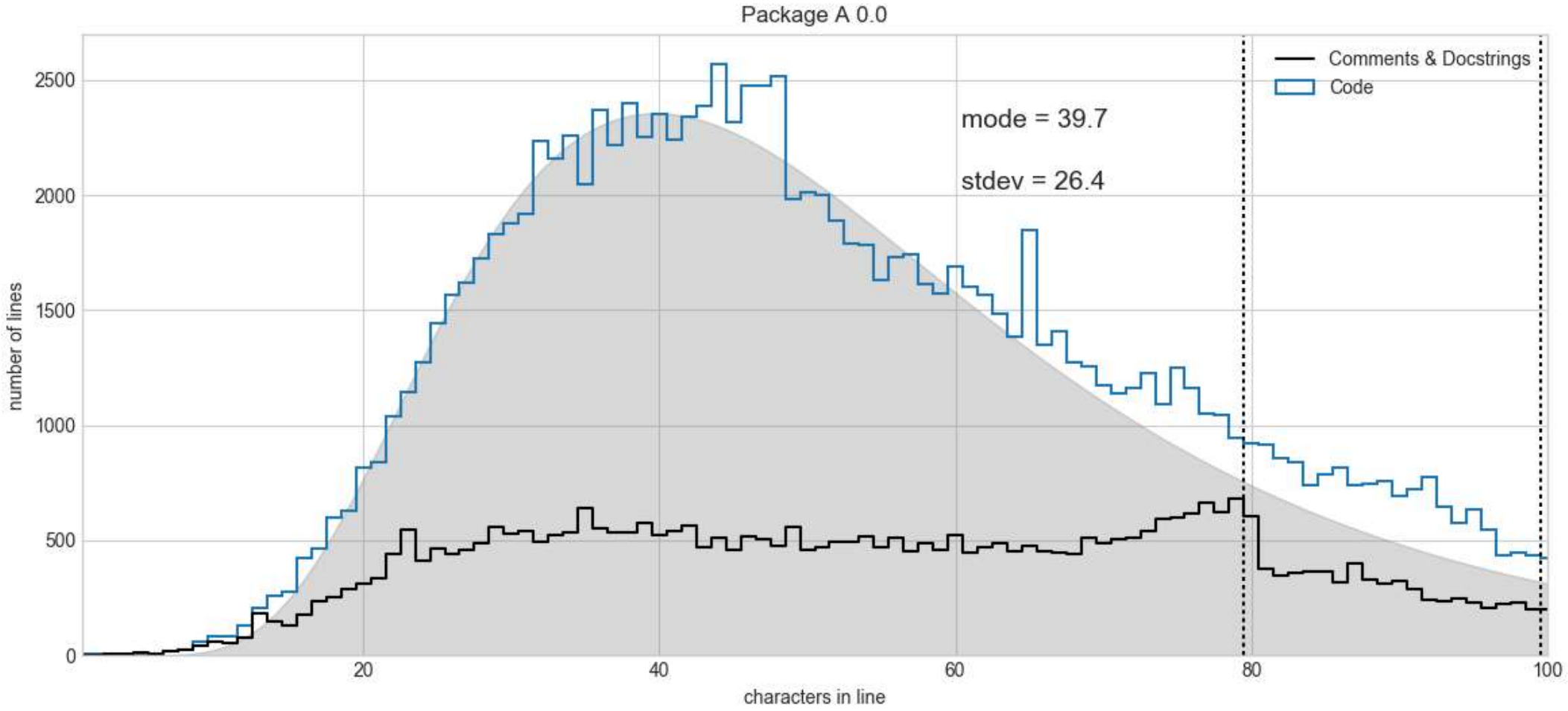
This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

CODE VS. COMMENTS & DOCSTRINGS OVERVIEW



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

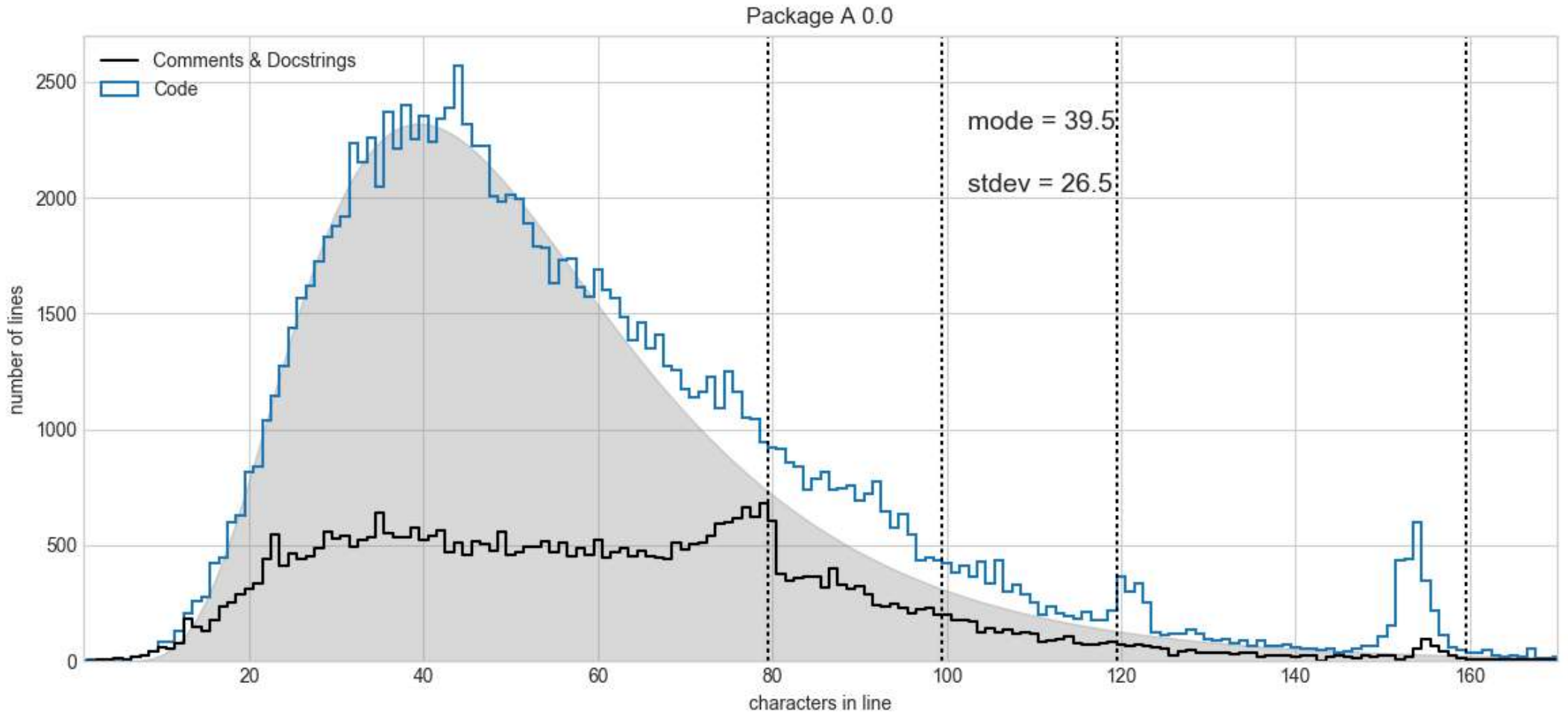
LET'S HAVE A LOOK ON CUSTOM CODE



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

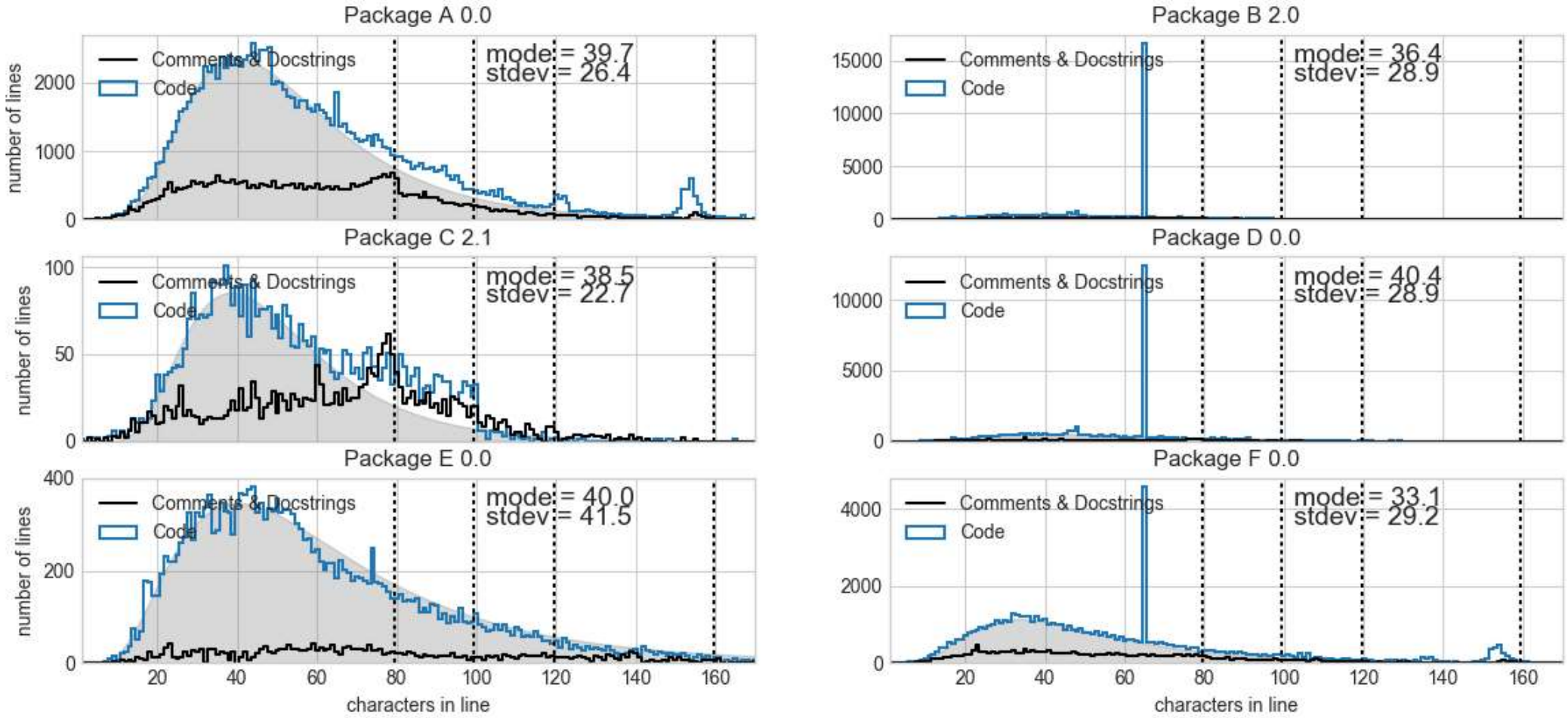
PUSH THE LIMITS
LINE LENGTH LIMITS

ROSEN
empowered by technology



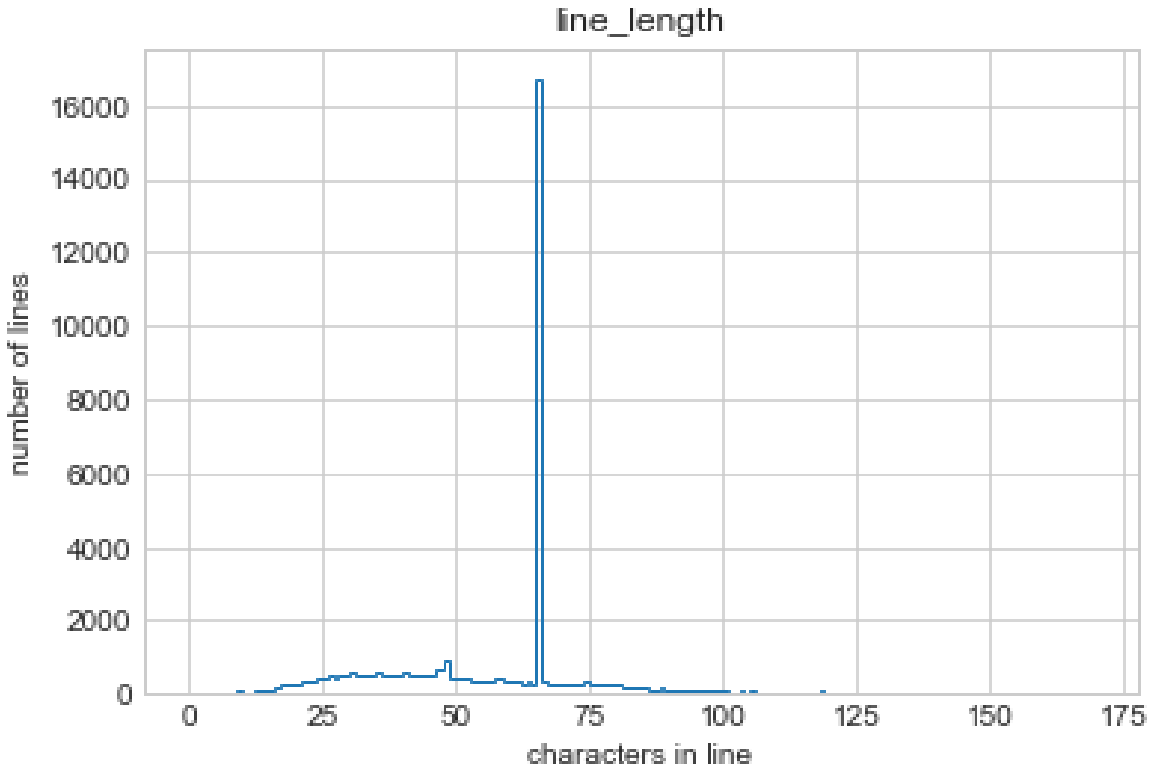
This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

MORE CUSTOM CODE OVERVIEW



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

ANALYZE THE UNUSUAL SPIKES
A CLOSER LOOK



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

ANALYZE THE UNUSUAL SPIKES
PANDAS TO THE RESCUE

ROSEN
empowered by technology

In [14]:
```
df = pd.DataFrame(data=[(len(line.rstrip('\n')), line) for line in set(iter_lines(alde_core))],
                  columns=('line_length', 'line'))

axis = df.hist(bins=range(170), histtype='step')
axis[0][0].set_xlabel("characters in line")
axis[0][0].set_ylabel("number of lines")
```

In [15]:
```
most_frequent_line_length = df['line_length'].value_counts().idxmax()
most_frequent_line_length
```

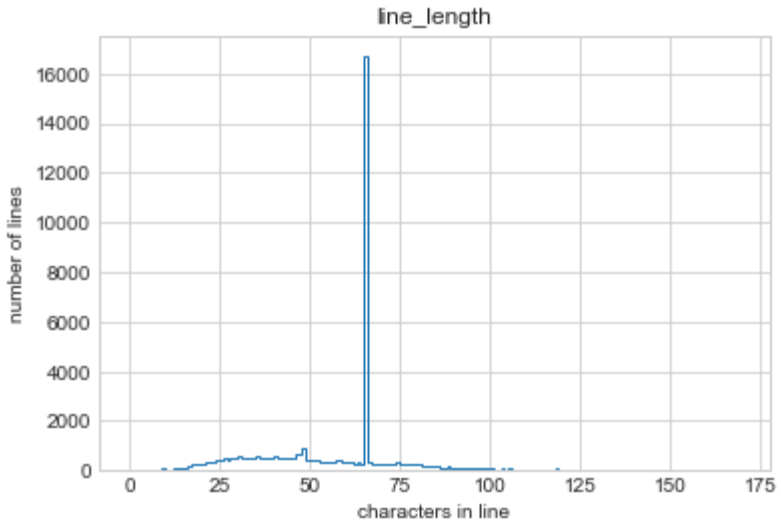
In [16]:
```
mask = df['line_length'] == most_frequent_line_length
df[mask].head(n=10)
```

Out[16]:

	line_length	line
2	65	\x69\xfc\xc0\x9d\x61\x63\xfd\x2e\xf9\x77\xf8\x...
3	65	\xfb\xd8\x84\xb8\xa8\xc7\x5f\x7c\xe9\x55\xc7\x...
4	65	\xef\x39\xd7\xb4\xf5\xce\x87\xc7\xcb\x65\x86\x...
7	65	parameter_model.chain.expert.max_samples_p...
8	65	\xef\xbe\x37\xb5\xd5\x9b\xb4\x6f\xde\xbd\xe7\x...
13	65	\x29\x1d\x79\x43\x91\x05\x85\x06\xa7\x3a\x6f\x...
18	65	\x9e\x93\x65\x59\xf7\xcf\x1a\x3b\x09\xbe\x09\x...
20	65	\x4d\x50\x54\x5c\x13\x37\xa0\xf4\x3b\x62\x8e\x...
23	65	\xa1\x76\x82\xcc\xf9\x2c\x13\xd2\x55\xff\xbb\x...
26	65	\x87\x4a\x2d\x64\x65\xd3\xe2\x67\x2a\x14\xe9\x...

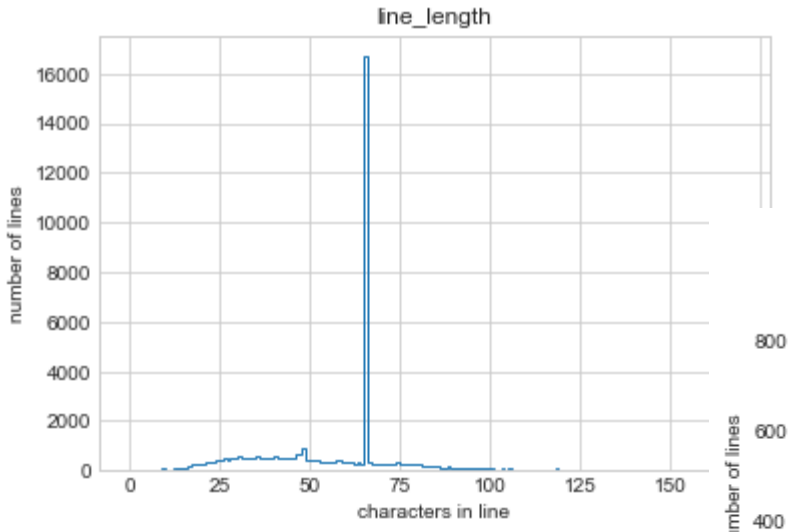
This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

ANALYZE THE UNUSUAL SPIKES
DATA CLEANSING

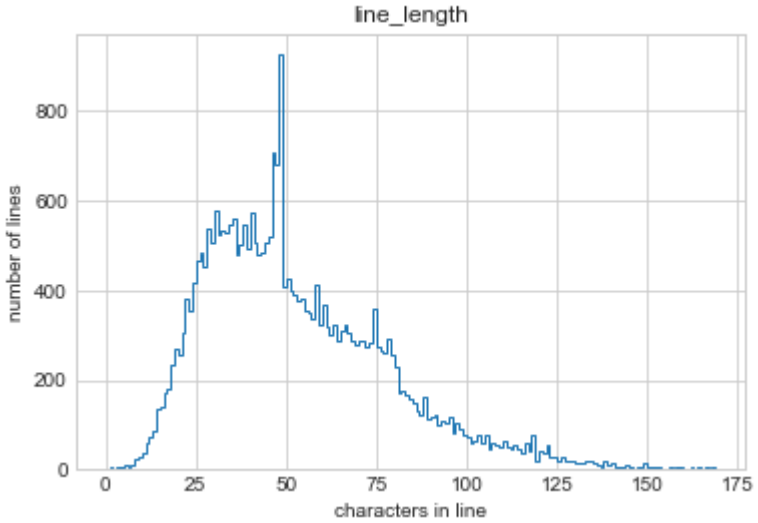


This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

ANALYZE THE UNUSUAL SPIKES
DATA CLEANSING

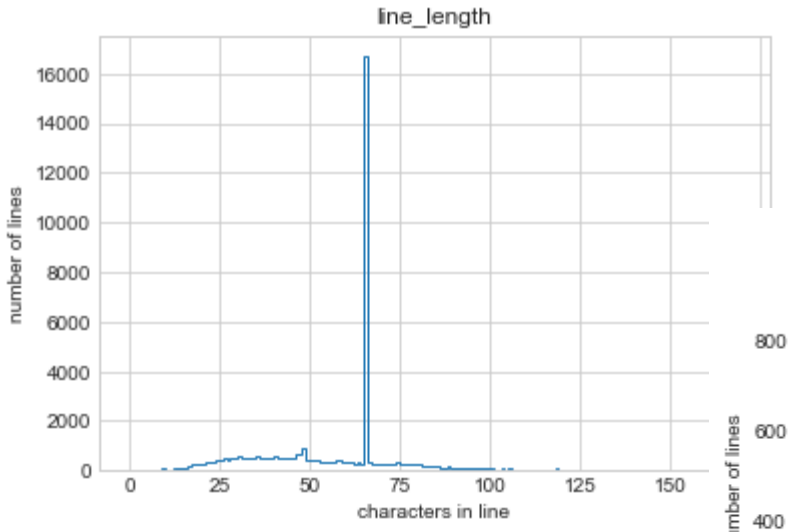


Postscript Icons

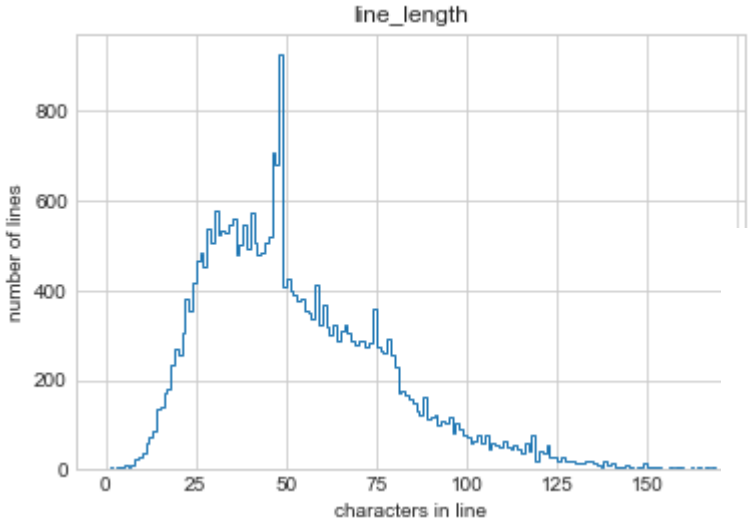


This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

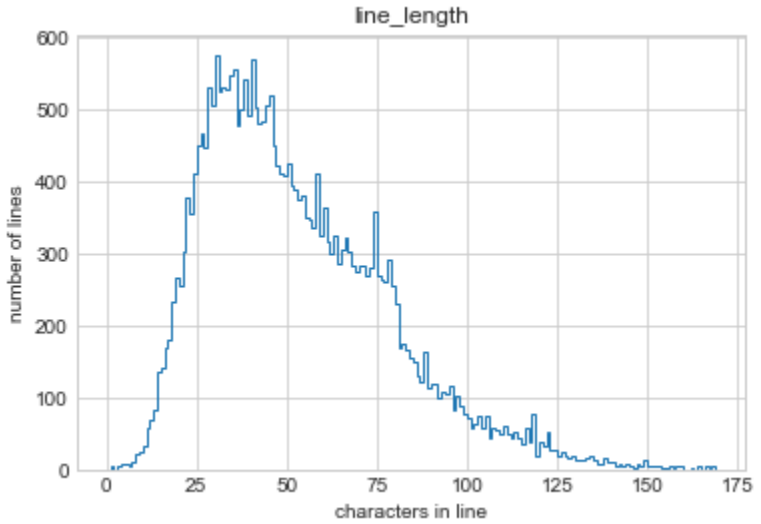
ANALYZE THE UNUSUAL SPIKES
DATA CLEANSING



Postscript Icons

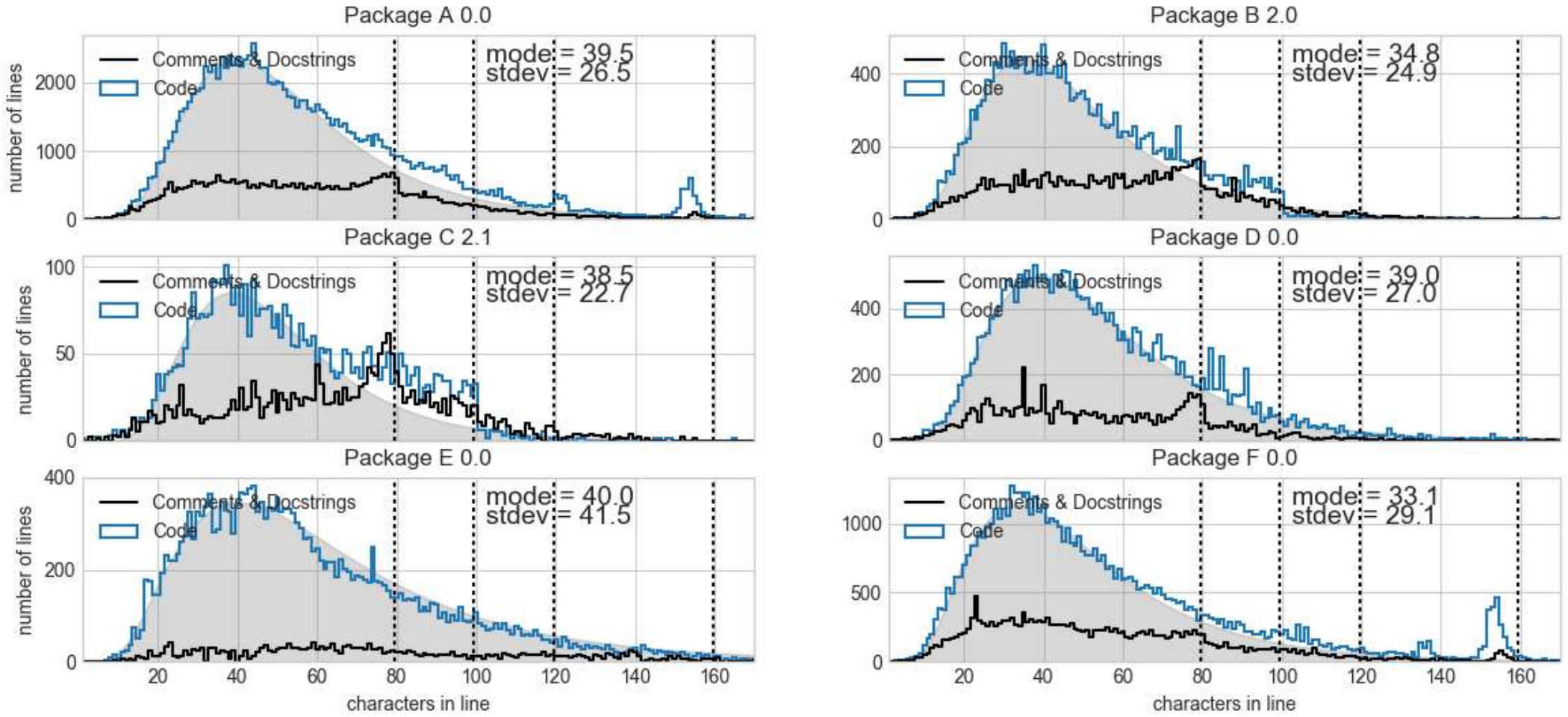


Colormaps



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

MORE CUSTOM CODE OVERVIEW PART II



This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

QUESTIONS

This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

RESOURCES

- http://jakevdp.github.io/blog/2017/11/09/exploring-line-lengths-in-python-packages/
- https://twitter.com/jakevdp/status/928759782625853440
- https://pandas.pydata.org/
- http://jupyter.org/

This document is the property of ROSEN Swiss AG who will safeguard its rights according to the civil and penal provisions of law. No part of this document may be reproduced or disclosed to any other party without the prior permission of ROSEN.

THANK YOU FOR JOINING THIS PRESENTATION.

ROSEN
empowered by technology