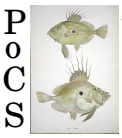


Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 300, 303, & 394
University of Vermont, Fall 2022
Assignment 20
"Permission to yell in a bloodcurdling way, sir?"

Due: Friday, March 3, by 11:59 pm
https://pdodds.w3.uvm.edu/teaching/courses/2022-2023pocsverse/assignments/20/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Assistant Deliverator: Dylan Casey (contact through Teams)
Office: The Ether
Office hours: See Teams calendar
Course website: https://pdodds.w3.uvm.edu/teaching/courses/2022-2023pocsverse
Overleaf: LaTeX templates and settings for all assignments are available at https://www.overleaf.com/project/631238b0281a33de67fc1c2b.

All parts are worth 3 points unless marked otherwise. Please show all your workingses clearly and list the names of others with whom you ~~conspired~~ collaborated.

For coding, we recommend you improve your skills with Python, R, and/or Julia. The (evil) Deliverator uses (evil) Matlab.

Graduate students are requested to use LaTeX (or related TeX variant). If you are new to LaTeX, please endeavor to submit at least n questions per assignment in LaTeX, where n is the assignment number.

Assignment submission:

Via Blackboard.

Bonus March facts:

The 0th day of March (or the last day of February) is always a Doomsday.

Pi Day is always a Doomsday.

For 2023: The Doomsday is Tuesday.

1. (3, 3, 3)

   Using the main text you chose at the start of the seasomester, plot happiness time series in the following ways using the labMT lexicon.

The labMT word list was published with Ref. [1] in 2011, and has been occasionally upgraded to accommodate major changes in language use.

See https://hedonometer.org and https://storywrangling.org for the current version.

(a) Process (destroy) your text so that it is a simple text file with one 1-gram per line—a vector of 1-grams.

To the extent possible, keep punctuation in as separate 1-grams. Periods, commas, semicolons, em dashes, ellipses, ...

You can submit this is a separate file, but okay to just say you've succeeded.

(b) First use the full lexical lens provided by labMT.

Make a single figure containing a stacked set of 7 plots with text windows of size $T = [10^z]$ for $z = 1$, 1.5, 2, 2.5, 3, 3.5, and 4.0.

Stacked here means separated and stacked vertically, as opposed to directly overlaid. See examples for Moby Dick at the end of this assignment.

The notation $[\cdot]$ means round to the nearest integer.

(c) Choose a 'good' text window from above, and repeat the analysis with lenses which exclude the central words around the neutral point.

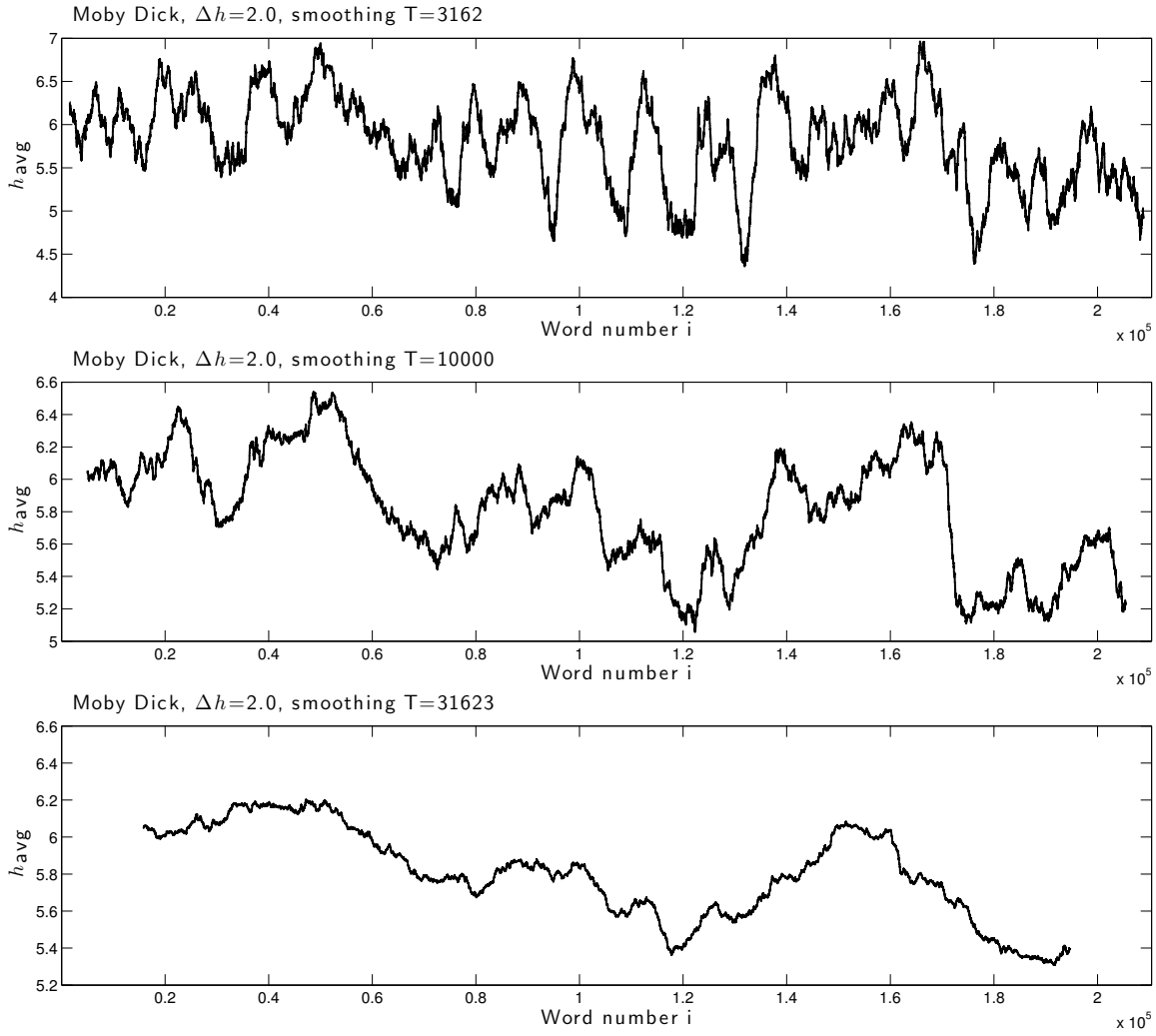
The blocked words are $h_{\text{avg}} \pm \delta h_{\text{avg}}$ where $\delta h_{\text{avg}} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,$ and 3.5.

By 'good', we mean one that seems to you to produce a reasonable smoothing. Not too choppy, not too washed out.

Notes:

- The horizontal axis is "reading-experience time" corresponding to 1-grams in the text, running from 1 to N.
- The windows should overlap, sliding one word ahead each time. This is a simple averaging filter.
- Points should be located above the center of each window.
- So the point for the window running from n to $n + T - 1$ (T words) will be located at $n + (T - 1)/2$.
- Do not pre-filter the text for any given lens. Windows will contain variable numbers of words with and without happiness scores.

Three example averaging windows for Moby Dick with $\delta h_{\text{avg}} = 2.0$:



References

[1] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6:e26752, 2011. pdf