

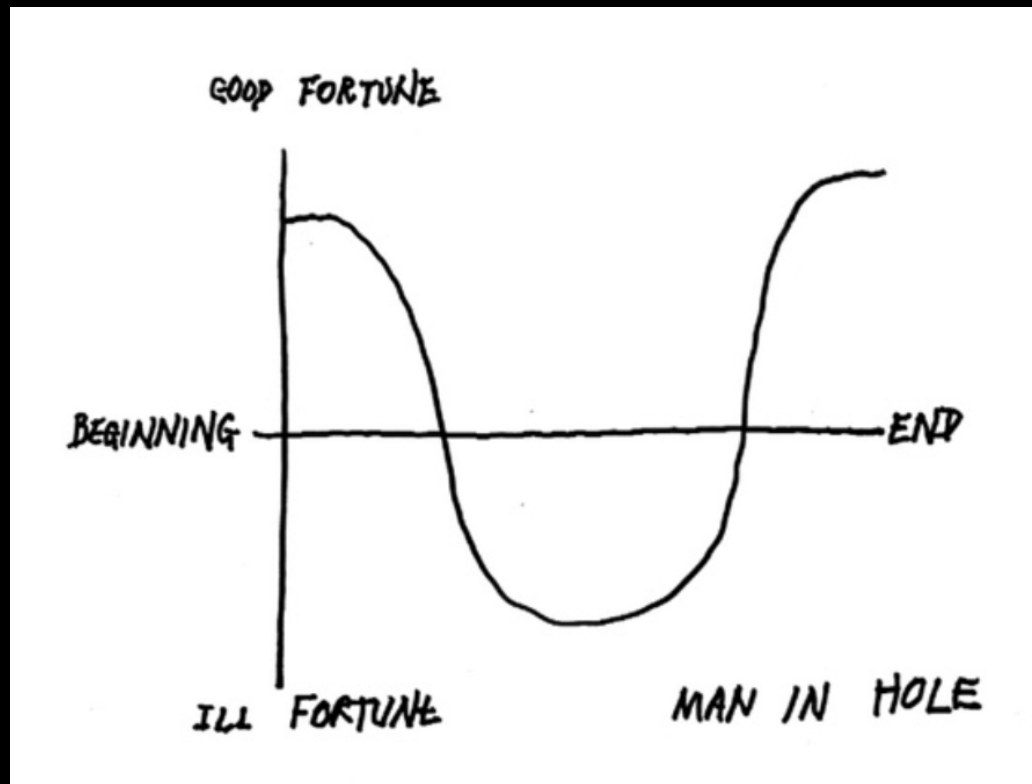
Sentiment Analysis as Proxy for Plot

Matthew L. Jockers
University of Nebraska
@mljockers

<https://www.youtube.com/watch?v=oP3cIh8v2ZQ>



“The fundamental idea is that stories have shapes which can be drawn on graph paper, and that the shape of a given society’s stories is at least as interesting as the shape of its pots or spearheads.”



“There is no reason why the simple shape of stories cannot be fed into computers.”

Positive Sentiments

“Once upon a time and a very good time it was there was a moocow coming down along the road and this moocow that was coming down along the road met a nicens little boy named baby tuckoo.” (James Joyce, *Portrait*)

“I love you as the grass loves the dew, as the birds love a bough.” (Dave Eggers, *The Circle*)

Negative Sentiments

“All the filth of the world, all the offal and scum of the world, we are told, shall run there as to a vast reeking sewer when the terrible conflagration of the last day has purged the world.” (James Joyce, *Portrait*)

“I mean, do you know the chaos this is wreaking on my family?” (Dave Eggers, *The Circle*)

Neutral Sentiments

She opened the door.

The hat was on the table.

I hated the way he looked at me that morning, and I was glad that he had become my friend.

syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text

Extracts sentiment and sentiment-derived plot arcs from text using three sentiment dictionaries conveniently packaged for consumption by R users. Implemented dictionaries include "syuzhet" (default) developed in the Nebraska Literary Lab "afinn" developed by Finn {\AA}rup Nielsen, "bing" developed by Minqing Hu and Bing Liu, and "nrc" developed by Mohammad, Saif M. and Turney, Peter D. Applicable references are available in README.md and in the documentation for the "get_sentiment" function. The package also provides a hack for implementing Stanford's coreNLP sentiment parser. The package provides several methods for plot arc normalization.

Version: 1.0.0
Imports: [openNLP](#), [NLP](#), [zoo](#), [dtt](#), stats, graphics
Suggests: [knitr](#), [pander](#), [testthat](#) (≥ 0.9.1)
Published: 2016-04-28
Author: Matthew Jockers [aut, cre]
Maintainer: Matthew Jockers <mjockers at gmail.com>
License: [GPL-3](#)
URL: <https://github.com/mjockers/syuzhet>
NeedsCompilation: no
Citation: [syuzhet citation info](#)
Materials: [README](#)
CRAN checks: [syuzhet results](#)

Downloads:

Reference manual: [syuzhet.pdf](#)
Vignettes: [Introduction to the Syuzhet Package](#)
Package source: [syuzhet_1.0.0.tar.gz](#)
Windows binaries: r-devel: [syuzhet_1.0.0.zip](#), r-release: [syuzhet_1.0.0.zip](#), r-oldrel: [syuzhet_1.0.0.zip](#)
OS X Mavericks binaries: r-release: [syuzhet_1.0.0.tgz](#), r-oldrel: [syuzhet_1.0.0.tgz](#)
Old sources: [syuzhet archive](#)

mjockers / syuzhet

Unwatch

28

Star

150

Fork

33

Code

Issues 2

Pull requests 0

Wiki

Pulse

Graphs

Settings

An R package for the extraction of sentiment and sentiment-based plot arcs from text — Edit

64 commits

1 branch

2 releases

4 contributors

Branch: master

New pull request

New file

Upload files

Find file

HTTPS

https://github.com/mjockers/syuzhet

Download ZIP

mjockers rebuild for CRAN		Latest commit 4de4d68 19 days ago
R	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
inst	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
man	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
tests	Version 1 update. See NEWS for details	21 days ago
vignettes	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
.Rbuildignore	rebuild for CRAN	19 days ago
.gitignore	added ds_store to gitignore	a year ago
.travis.yml	Fixing travis.yml with java suggestion from benmarwick	a year ago
DESCRIPTION	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
NAMESPACE	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
NEWS	Version 1 update. See NEWS for details	21 days ago
README.md	Version 1 update. See NEWS for details	21 days ago
cran-comments.md	Minor tweaks to be compliant with CRAN changes effective June 2015. F...	19 days ago
syuzhet.Rproj	Initial Commit	a year ago

Localized Sentiment Flow

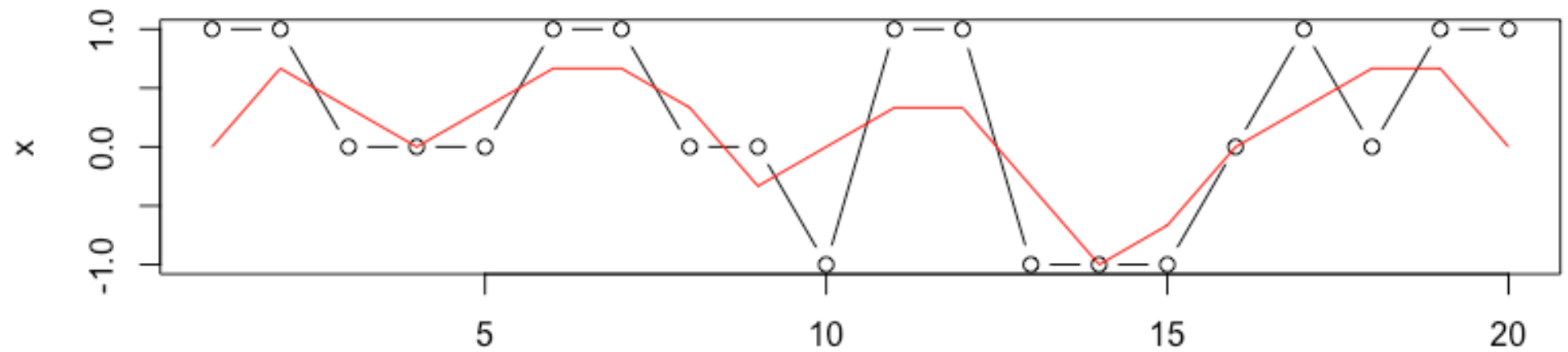
"I grew to hate him with all my heart." (-)

"He was not at all nice or generous." (+)

"I vowed that I would someday kill him." (-)

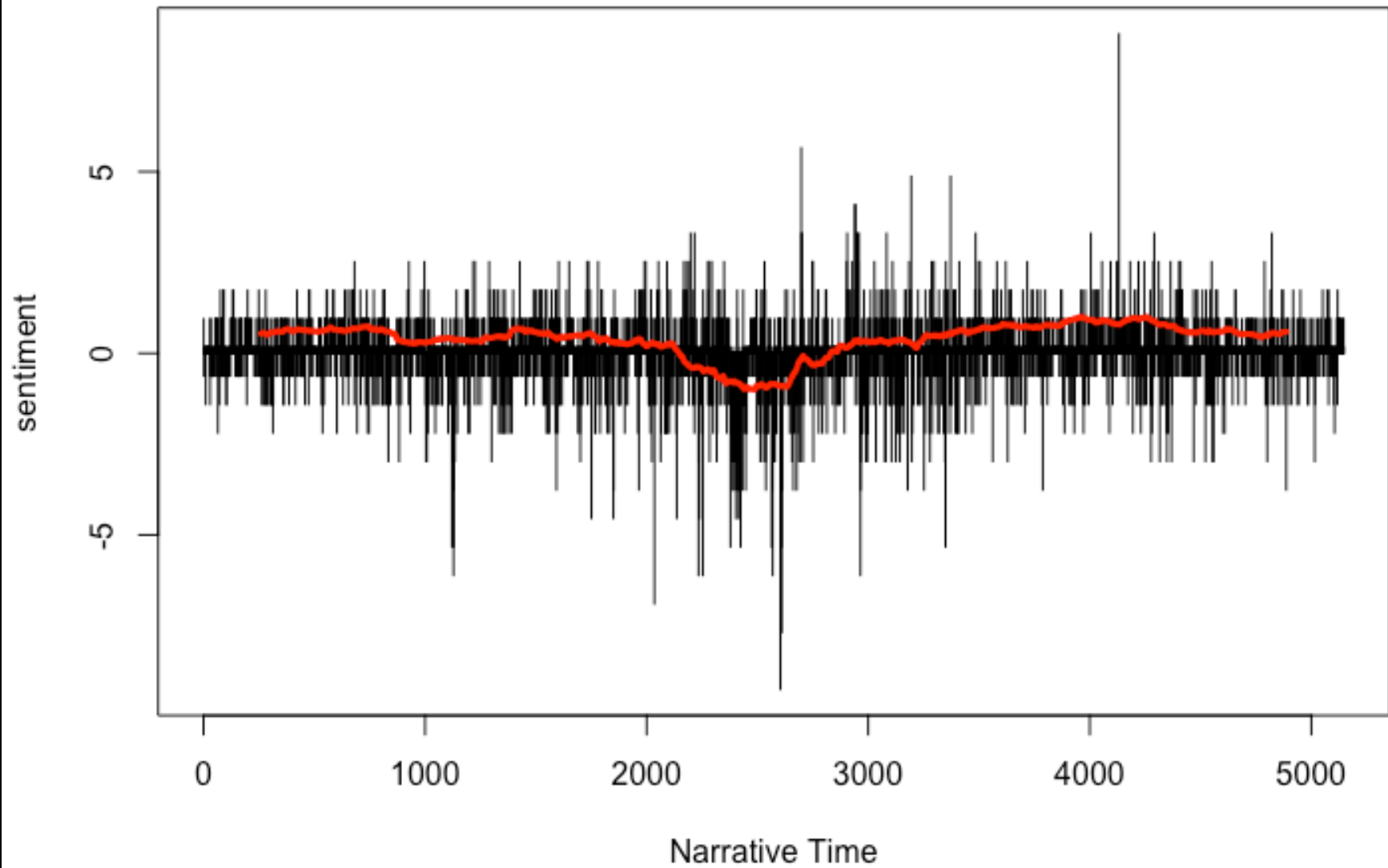
```
> library(syuzhet)
> sample <- "I grew to hate him with all my heart. He was not at all
nice or generous. I vowed that I would someday kill him."
> sent_v <- get_sentences(sample)
> result <- get_sentiment(sent_v, method = "afinn")
> mean(result)
[1] -0.3333333
```

1, 1, 0, 0, 0, 1, 1, 0, 0, -1, 1, 1, -1, -1, -1, 0, 1, 0, 1, 1



0, 0.7, 0.3, 0, 0.3, 0.7, 0.7, 0.3, -0.3, 0, 0.3, 0.3, -0.3, -1, -0.7, 0, 0.3, 0.7, 0.7, 0

Portrait of the Artist

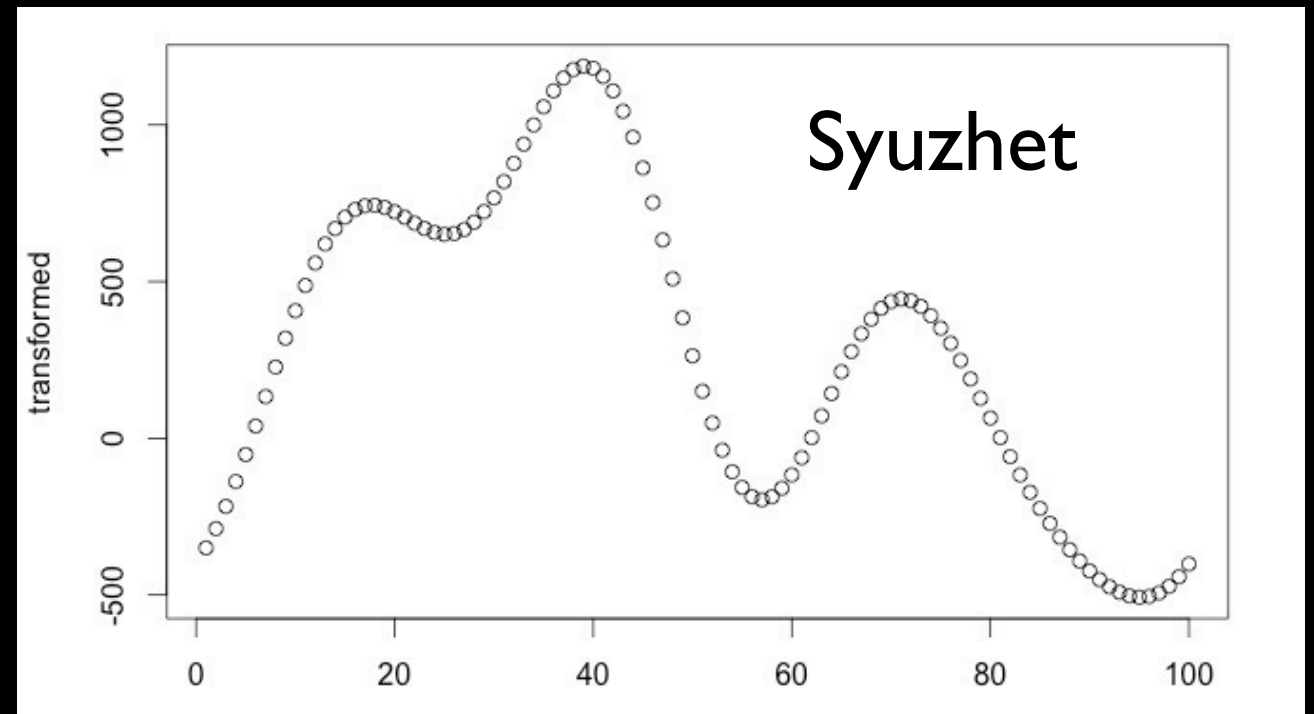
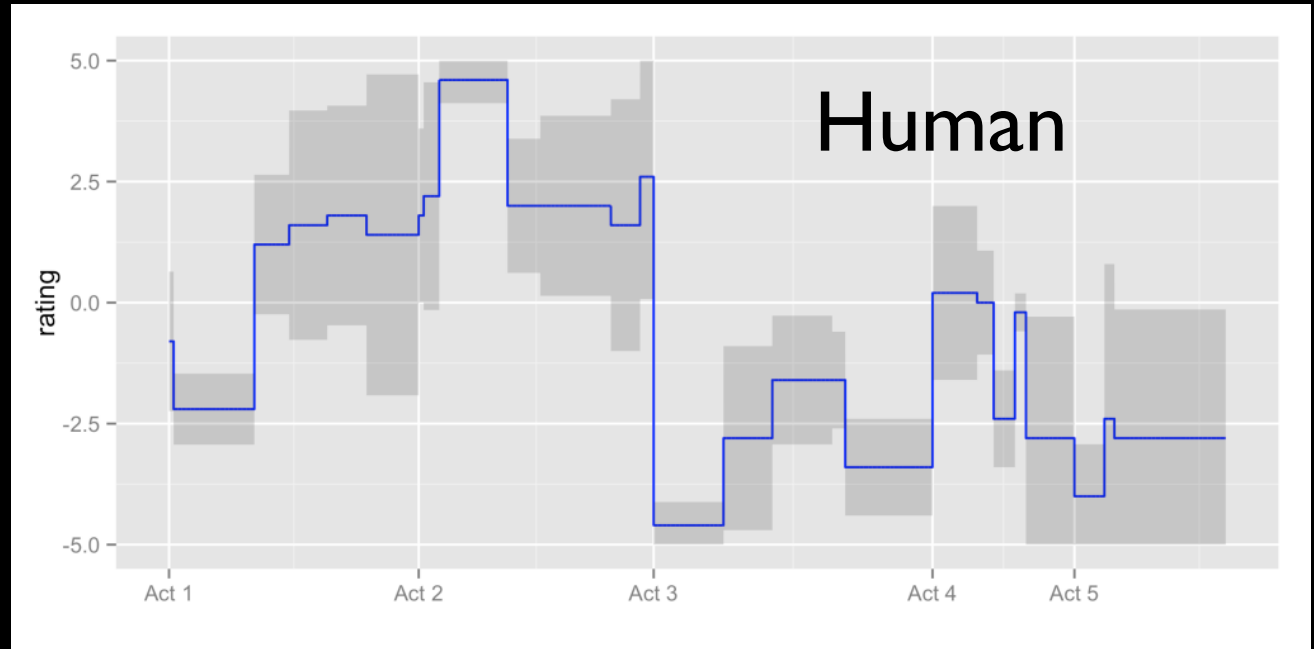


Portrait of the Artist as a Young Man



Mean Machine to Human Correlation Coefficient: 0.85
Polarity Agreement (Human to Machine) 73.49

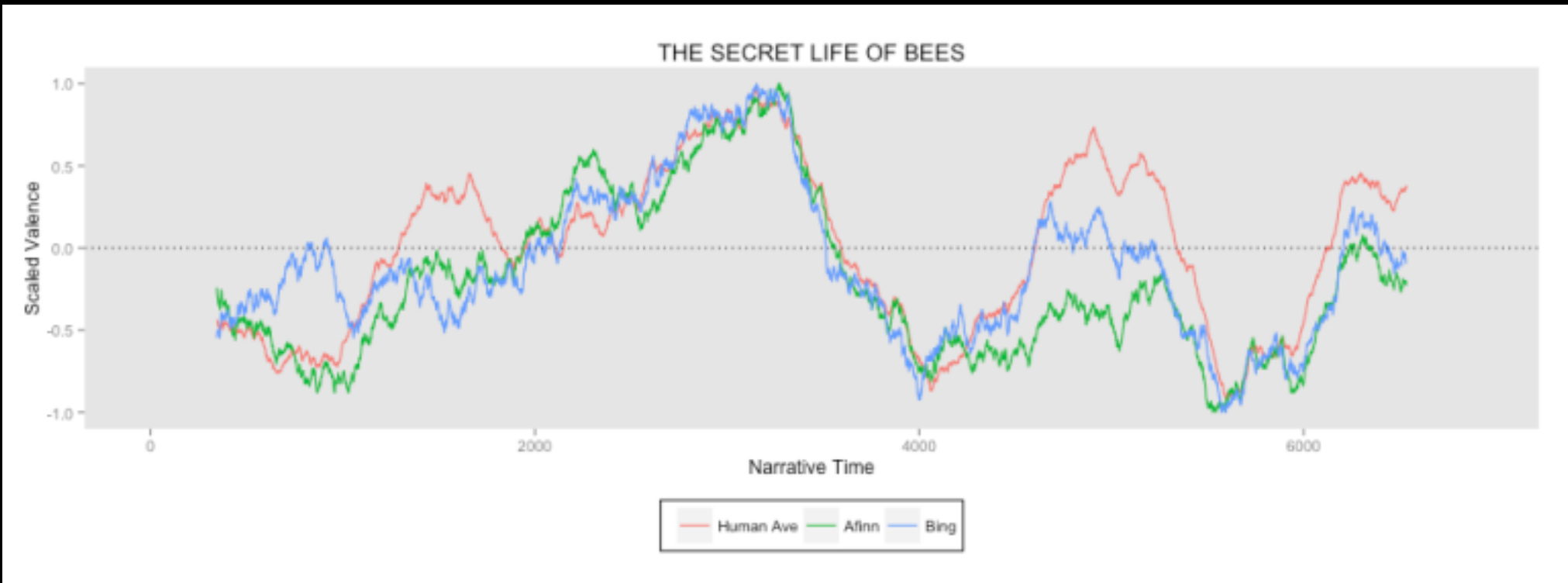
Romeo and Juliet



Images from Bamman
and Underwood

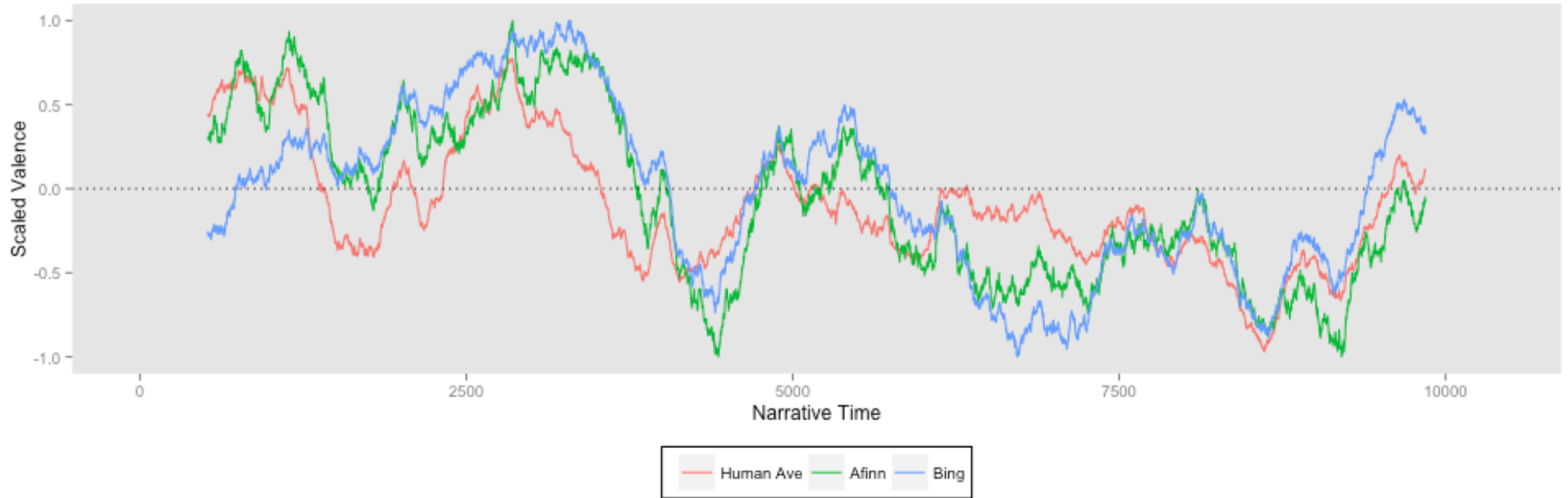


1. Mean Inter-Coder Correlation Coefficient: 0.87
2. Mean Machine to Mean Human Correlation Coefficient: 0.72
3. Human Machine Polar Disagreement 24%
4. Human Machine Polar Disagreement (allowing neutral) 11%

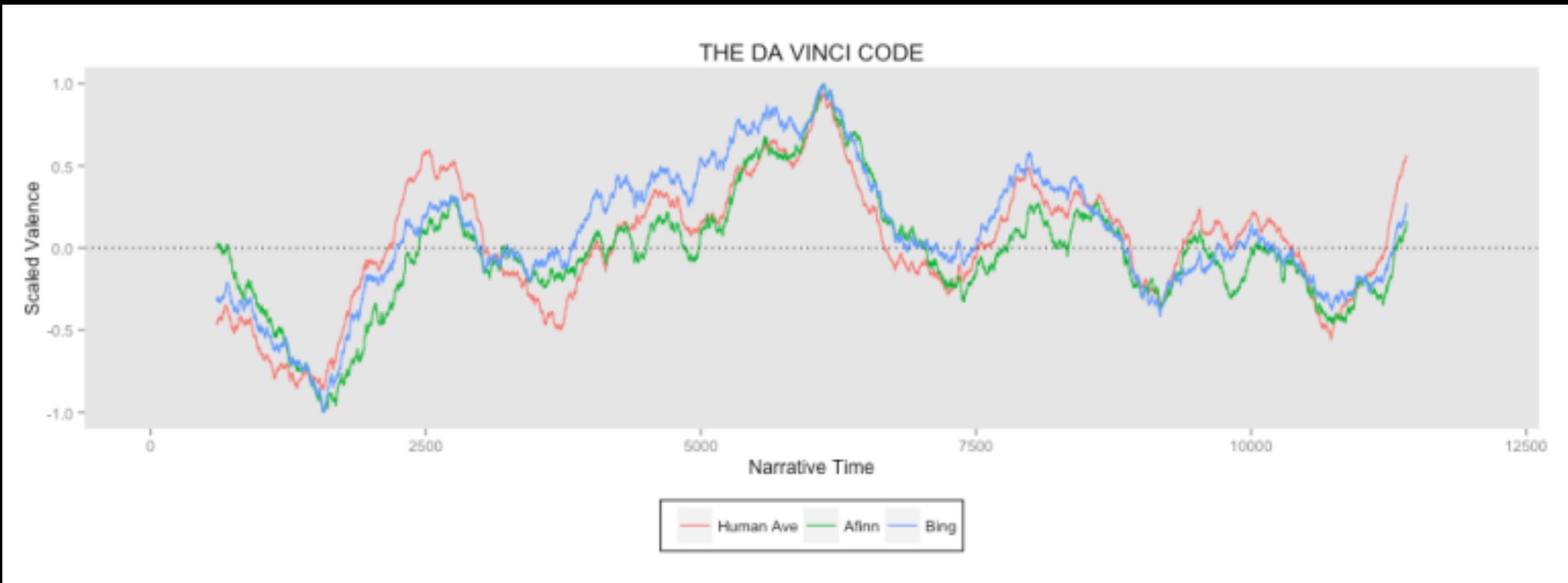


1. Mean Inter-Coder Correlation Coefficient: 0.9
2. Mean Machine to Mean Human Correlation Coefficient: 0.79
3. Human Machine Polar Disagreement 28%
4. Human Machine Polar Disagreement (allowing neutral) 13%

ALL THE LIGHT WE CANNOT SEE



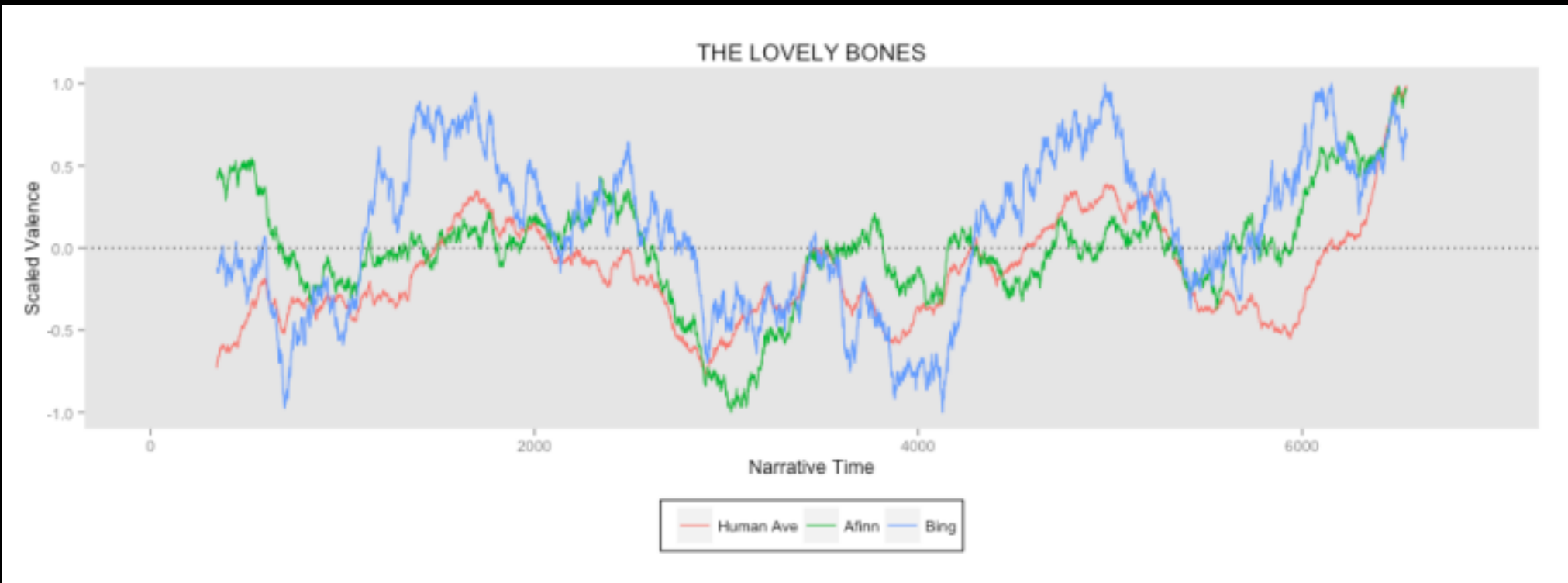
1. Mean Inter-Coder Correlation Coefficient: 0.73
2. Mean Machine to Mean Human Correlation Coefficient: 0.69
3. Human Machine Polar Disagreement 23%
4. Human Machine Polar Disagreement (allowing neutral) 8%



1. Mean Inter-Coder Correlation Coefficient: 0.68
2. Mean Machine to Mean Human Correlation Coefficient: 0.79
3. Human Machine Polar Disagreement 20%
4. Human Machine Polar Disagreement (allowing neutral) 9%

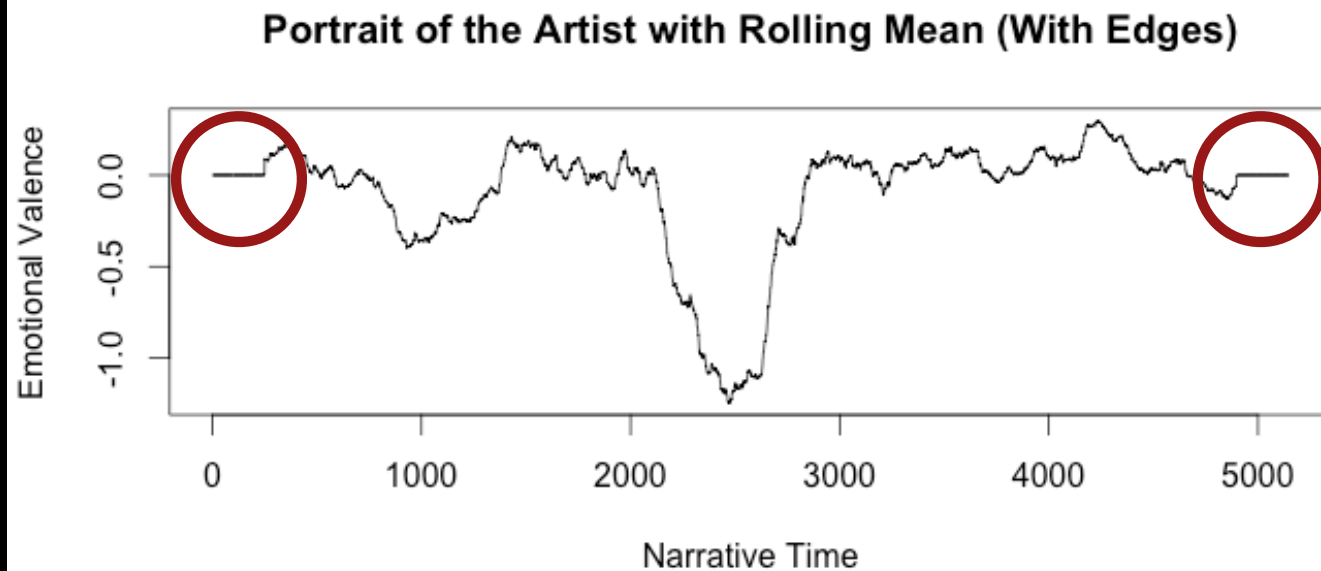
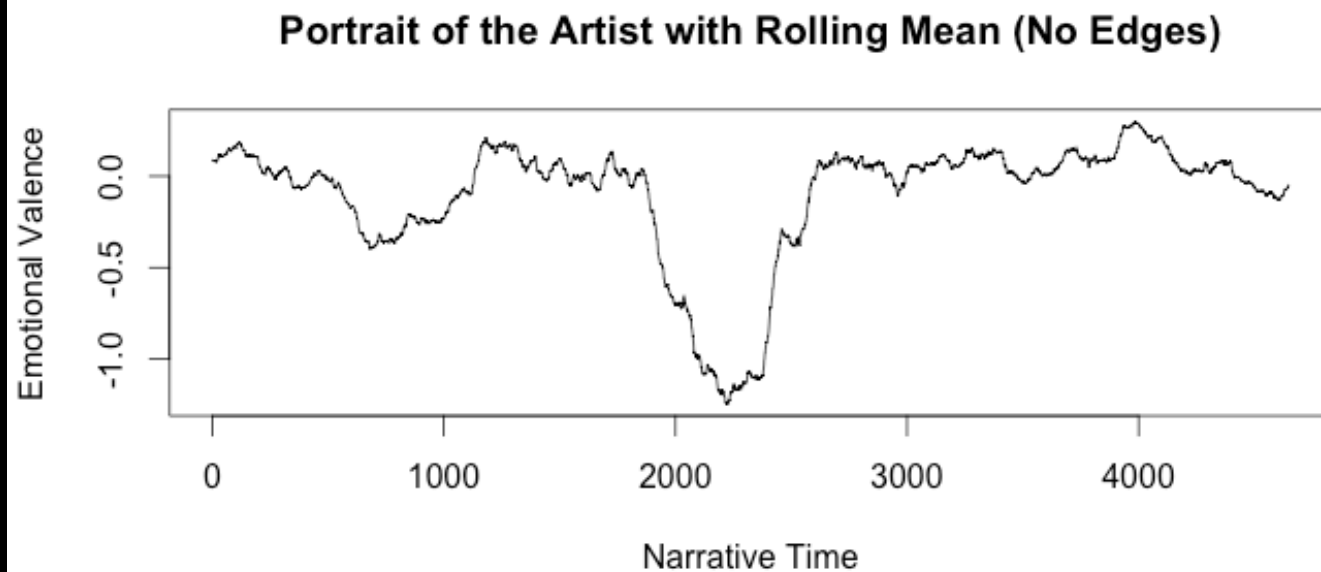


1. Mean Inter-Coder Correlation Coefficient: 0.65
2. Mean Machine to Mean Human Correlation Coefficient: 0.66
3. Human Machine Polar Disagreement 23%
4. Human Machine Polar Disagreement (allowing neutral) 11%

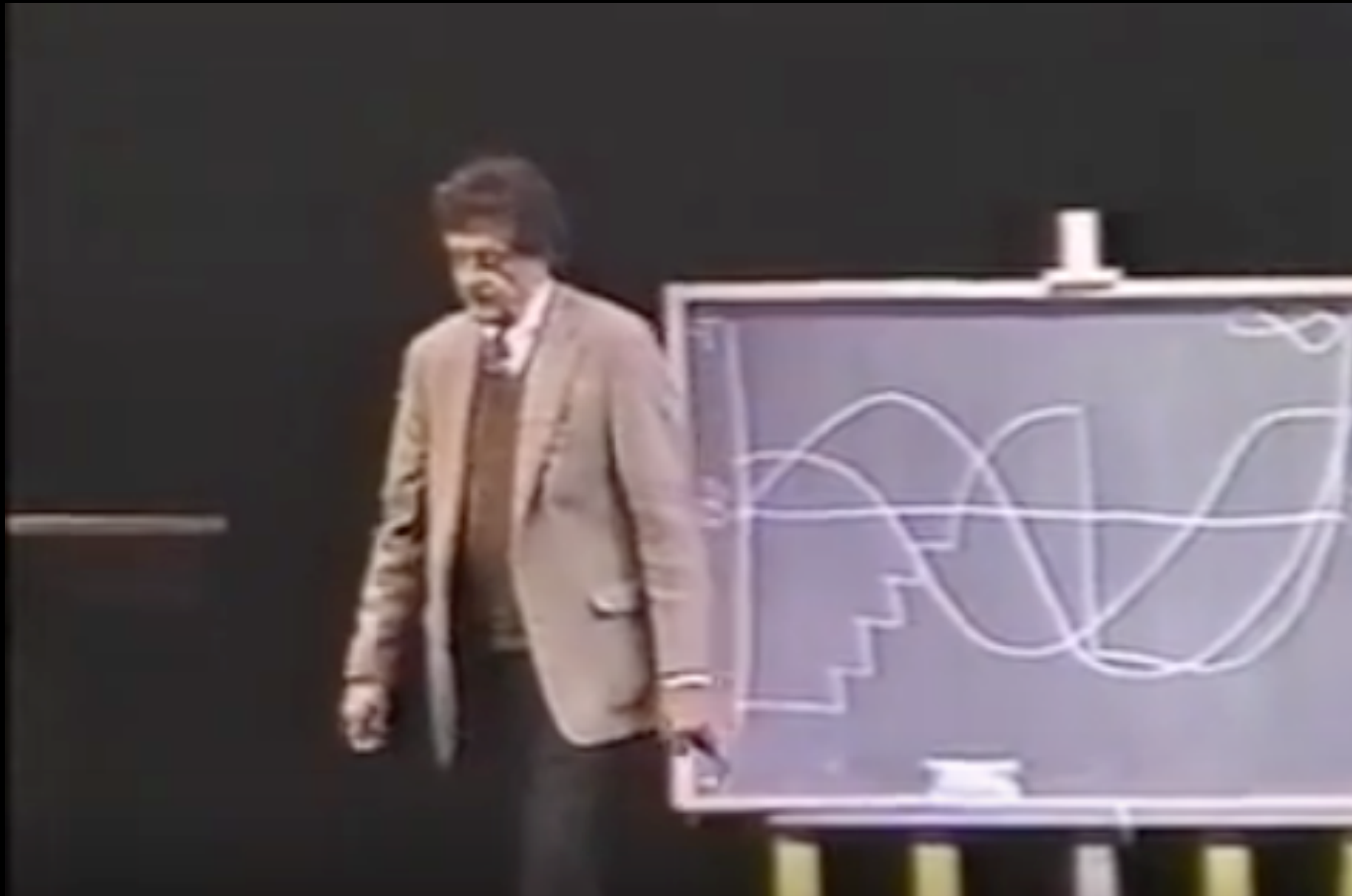


1. Mean Inter-Coder Correlation Coefficient: 0.62
2. Mean Machine to Mean Human Correlation Coefficient: 0.6
3. Human Machine Polar Disagreement 25%
4. Human Machine Polar Disagreement (allowing neutral) 11%

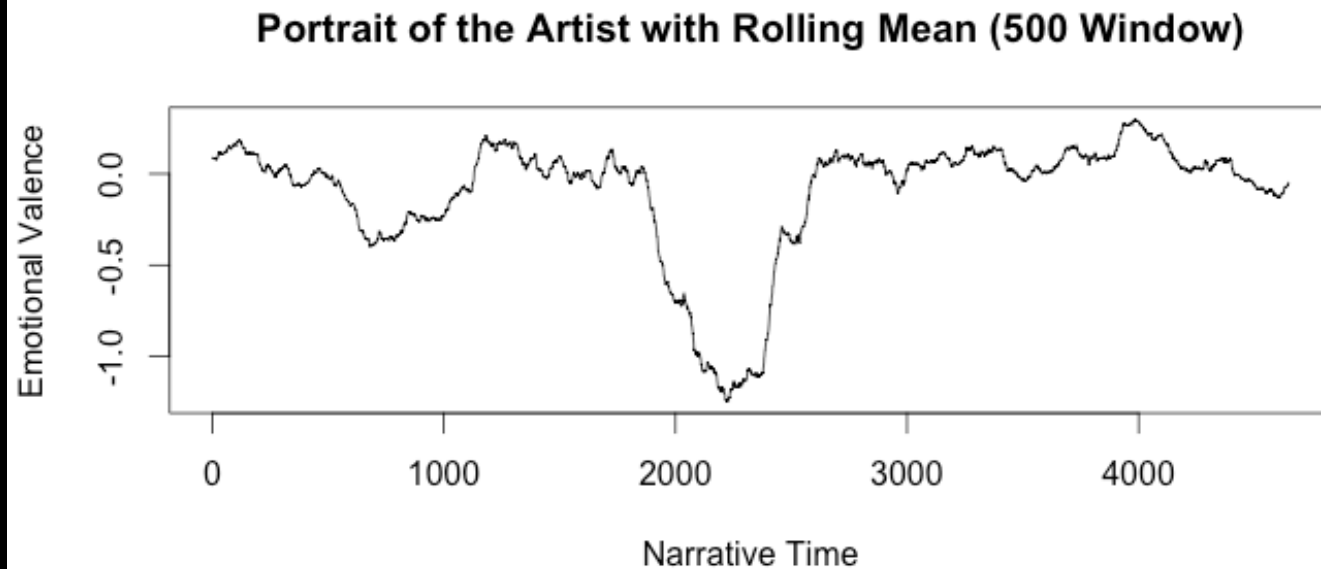
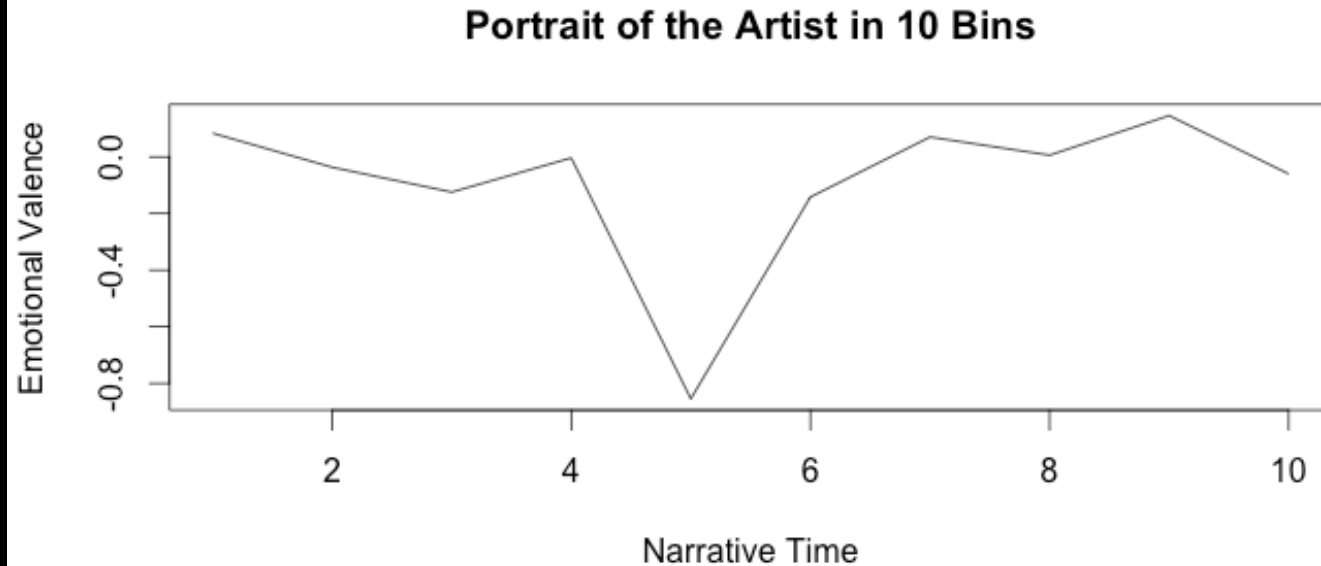
With rolling means we lose the edges



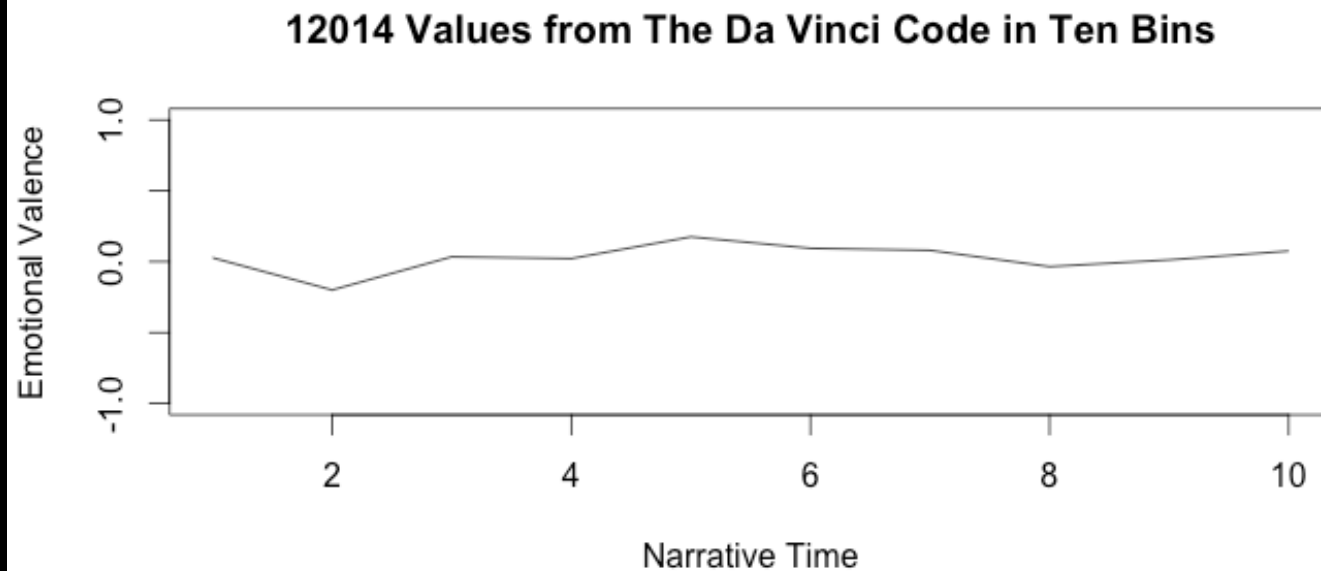
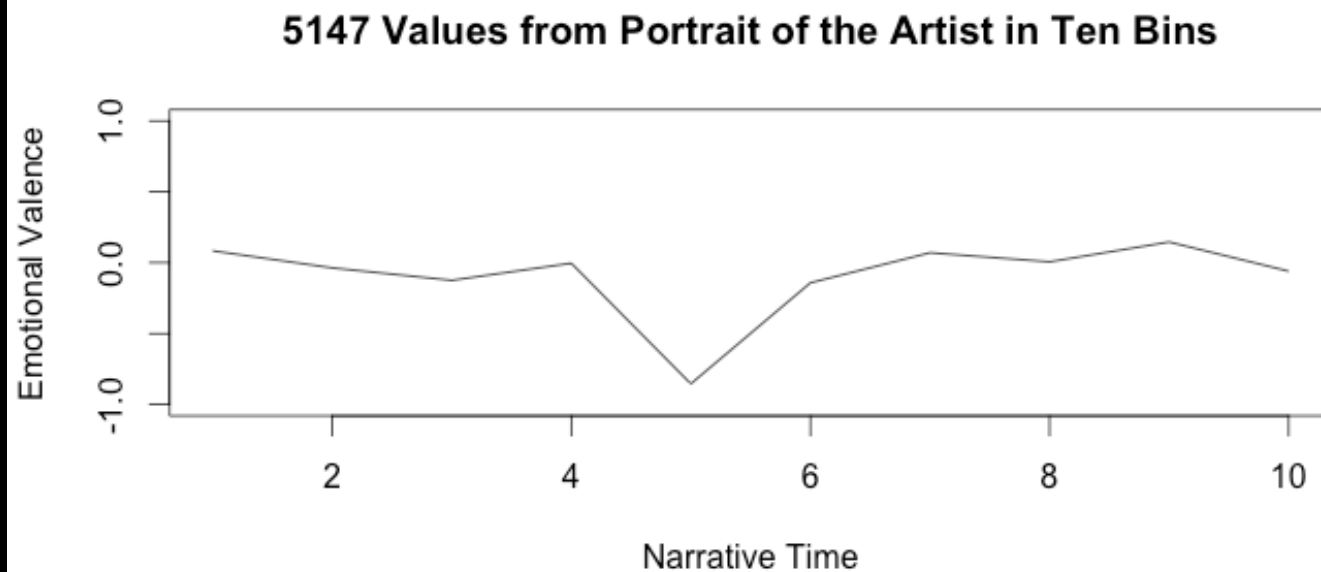
The B-E Axis Problem



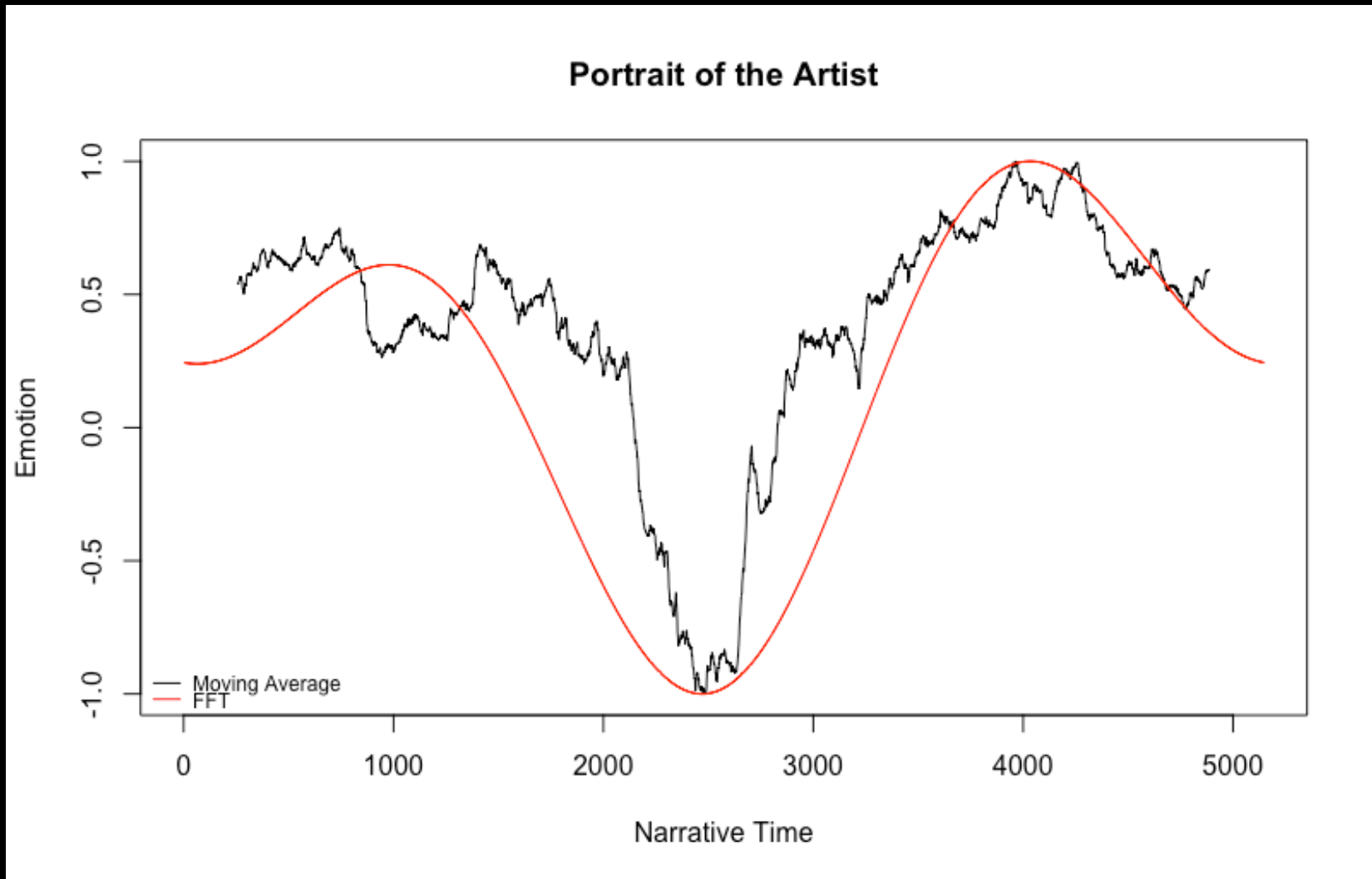
Binning (looks good here, but ...)



Binning not so good for comparison...

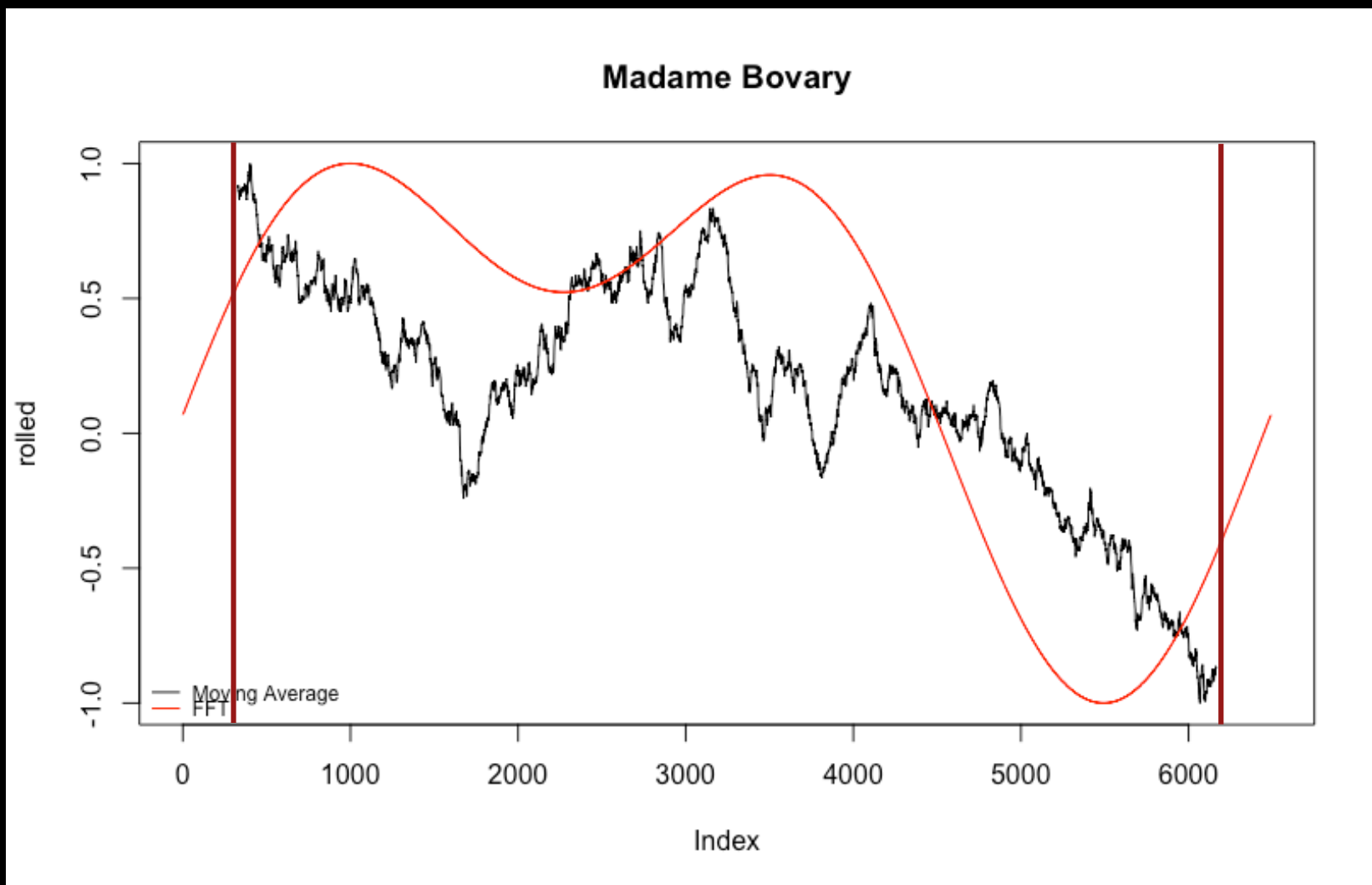


Solution 1.0



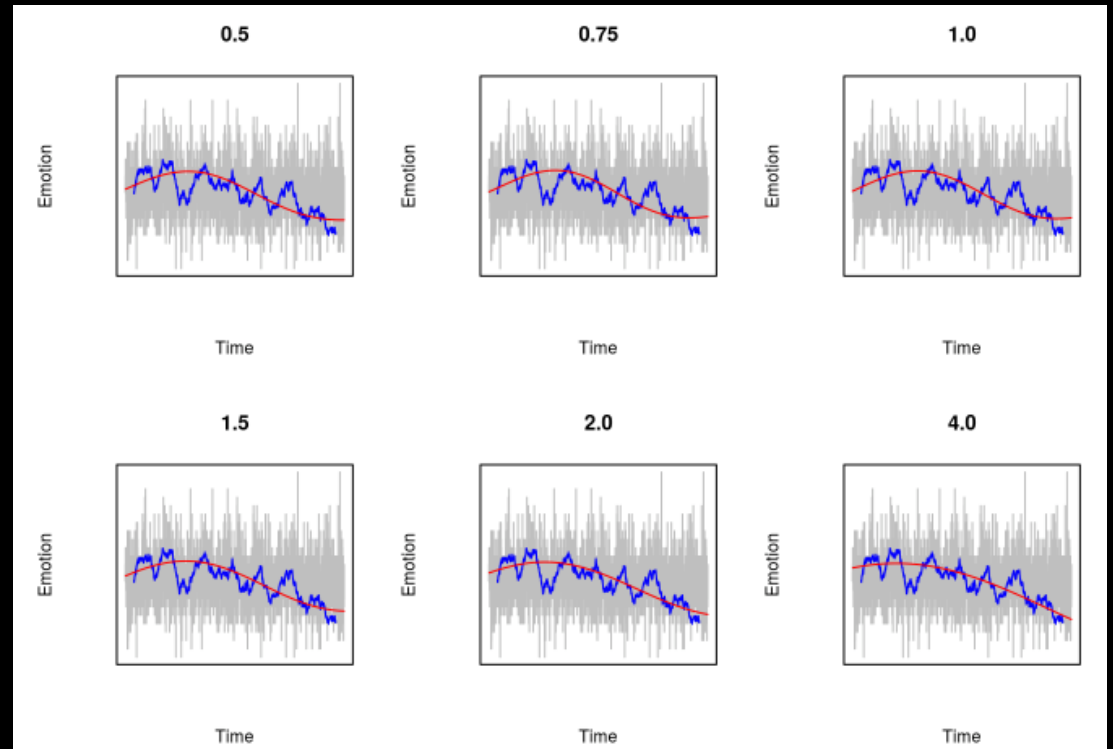
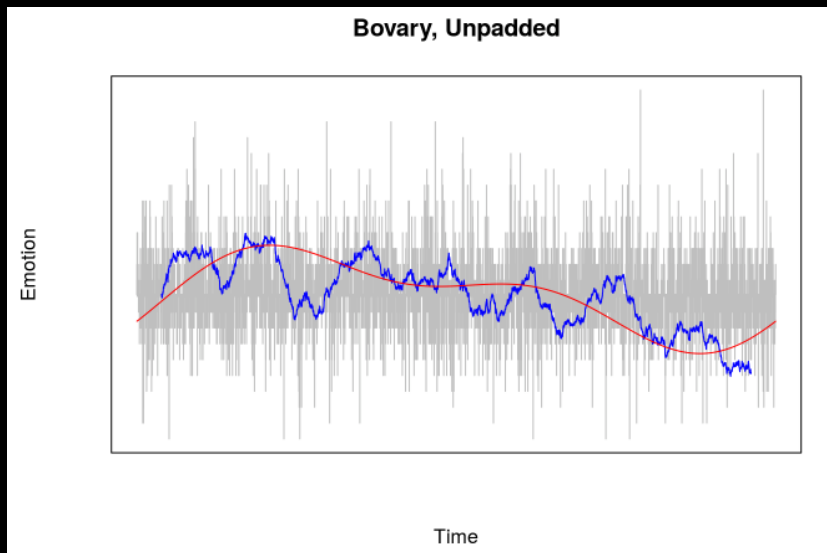
A Beautiful “Man in Hole”

Solution 1.0



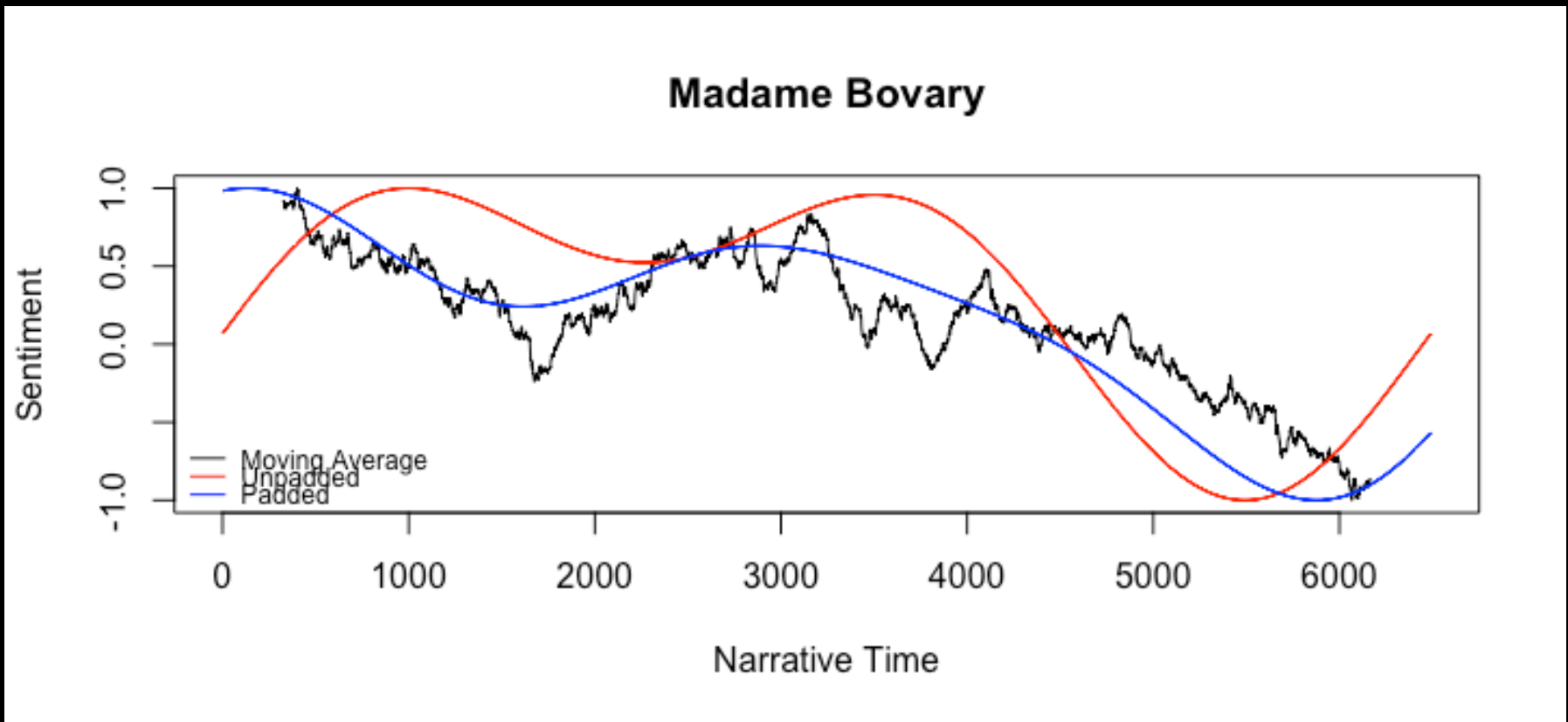
Bad Edges, Ugh!

Solution 1.1



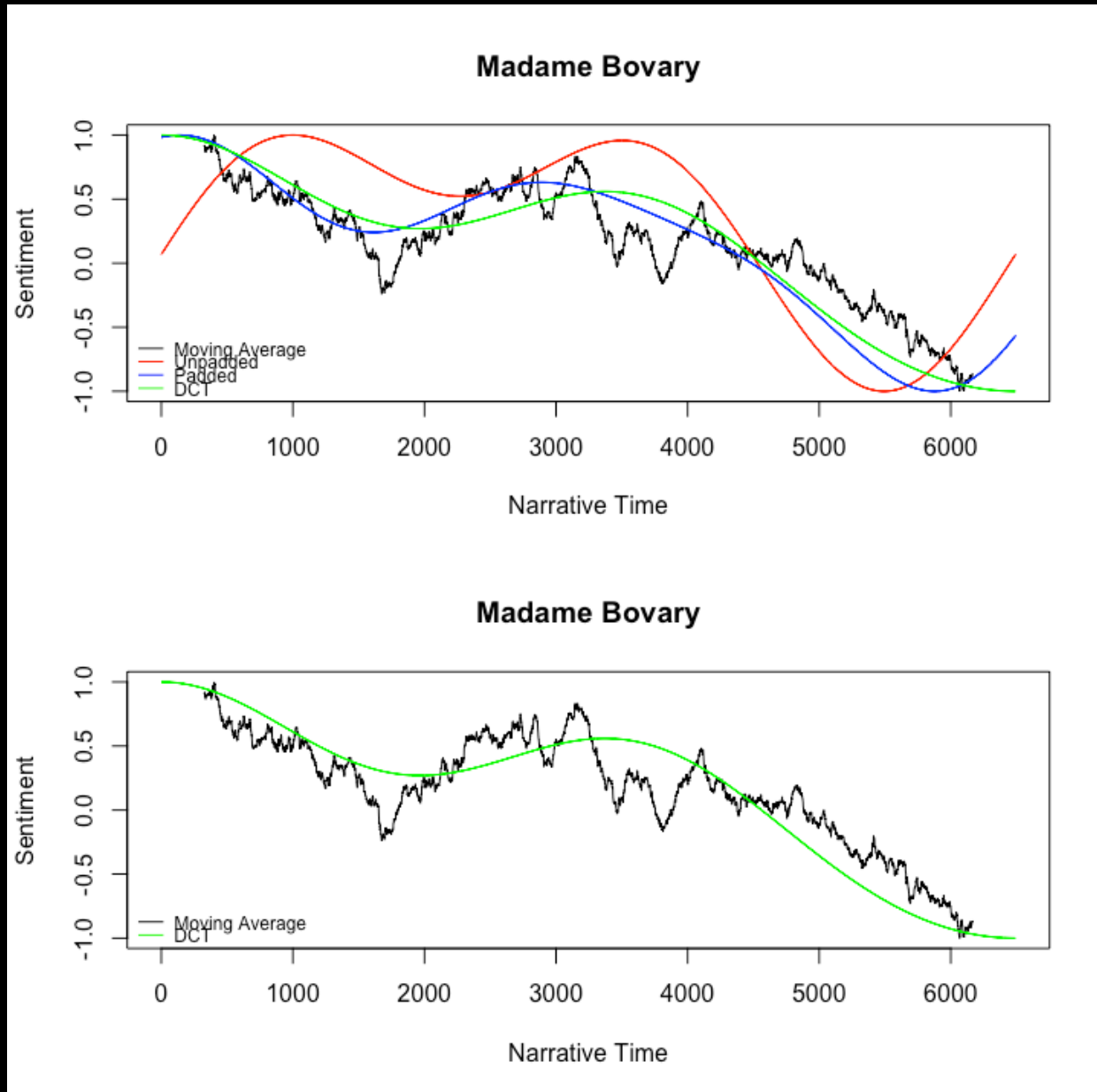
Images from Tommy M. McGuire http://www.crsr.net/files/Exploring_Syuzhet.html

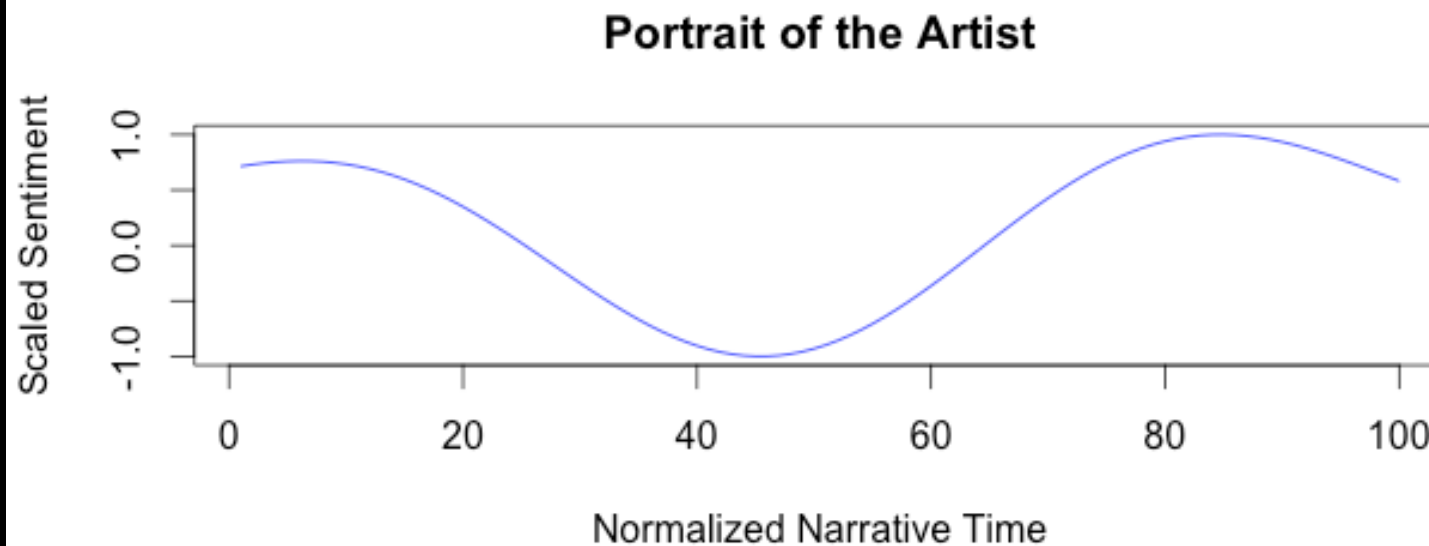
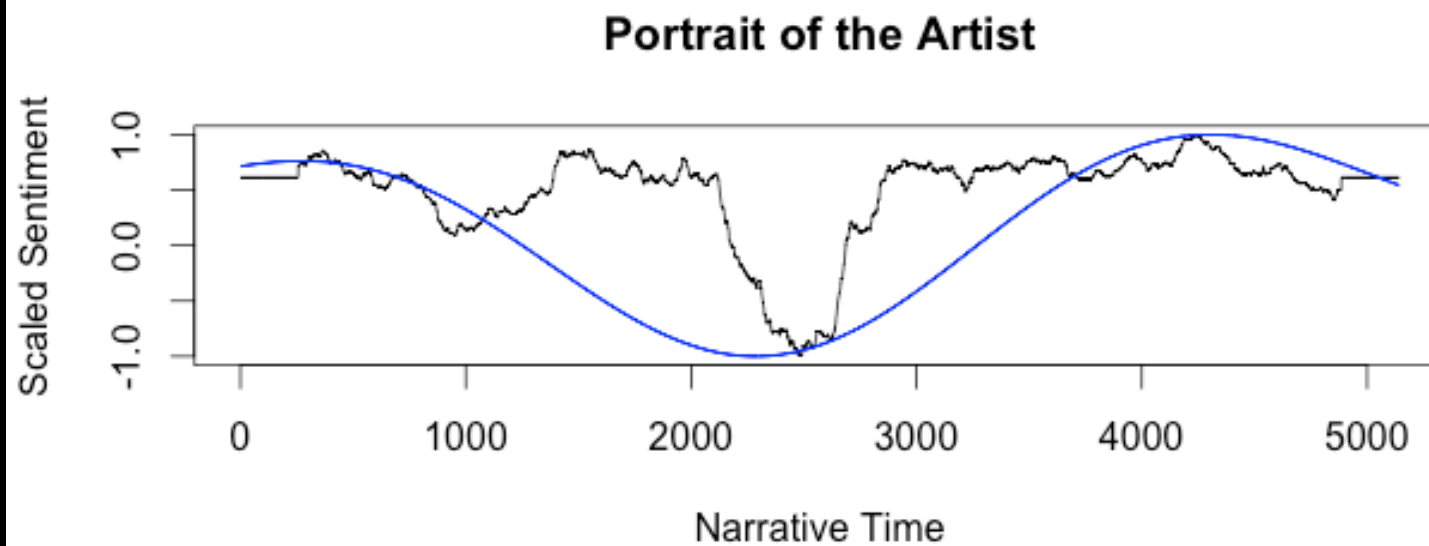
Solution 1.1

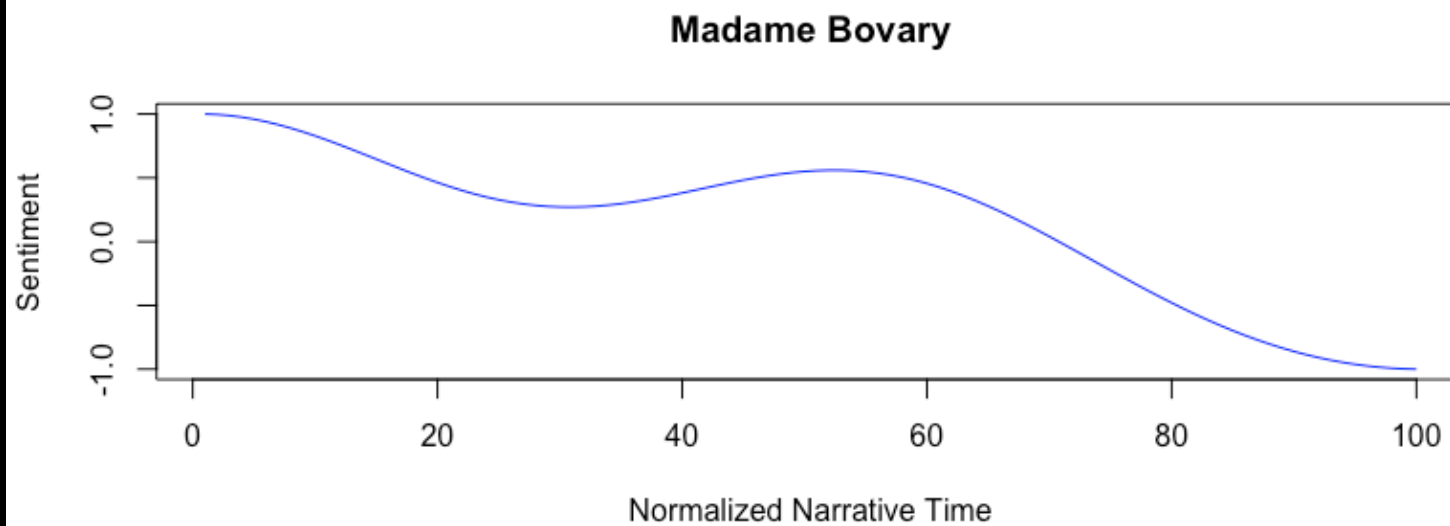
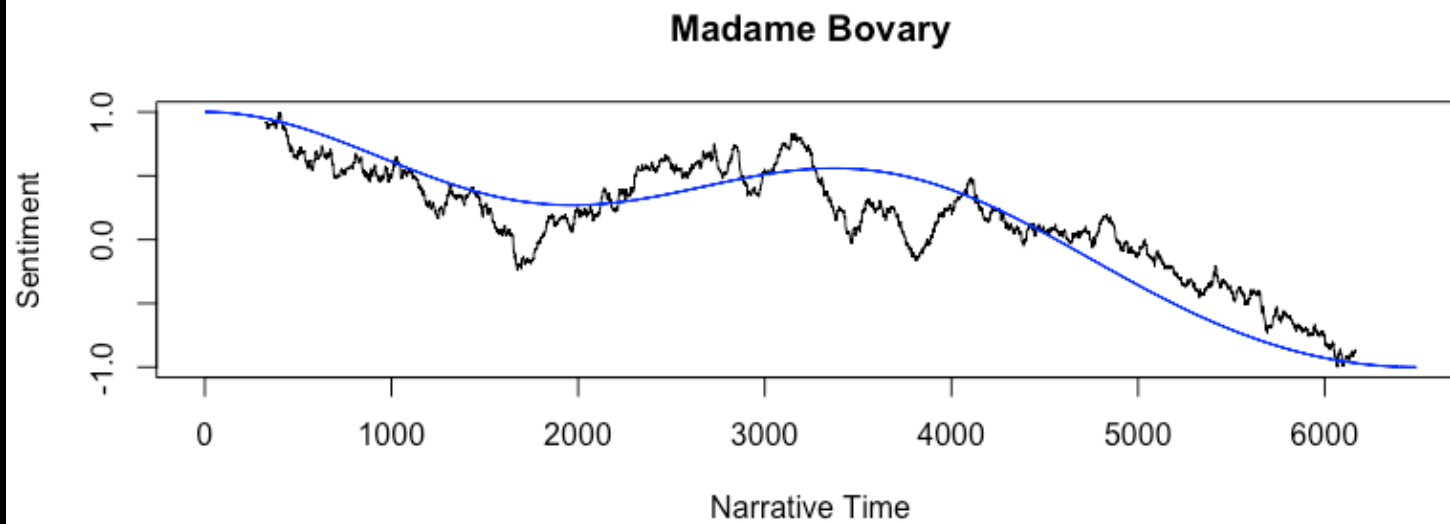


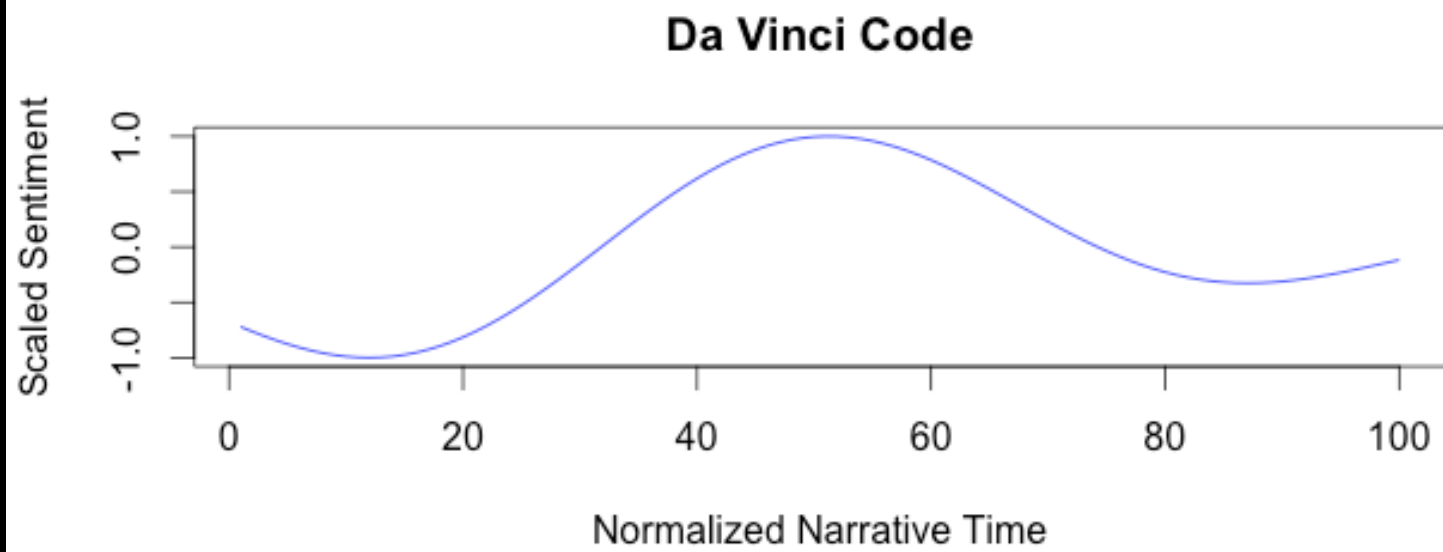
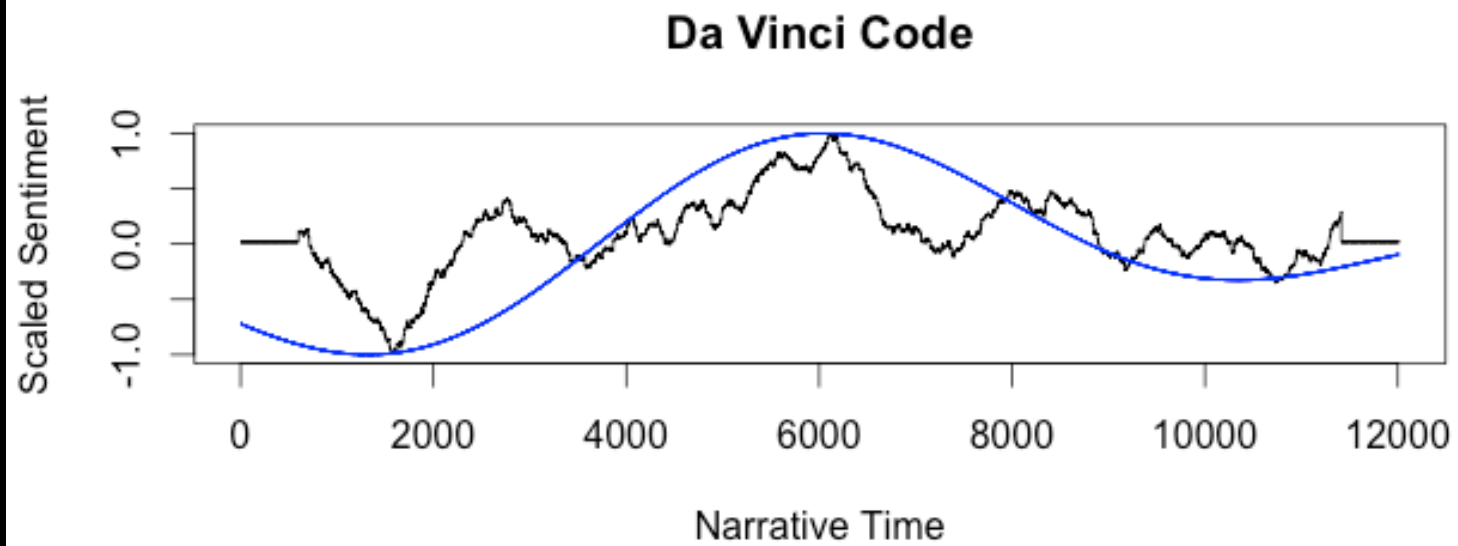
Solution 2.0

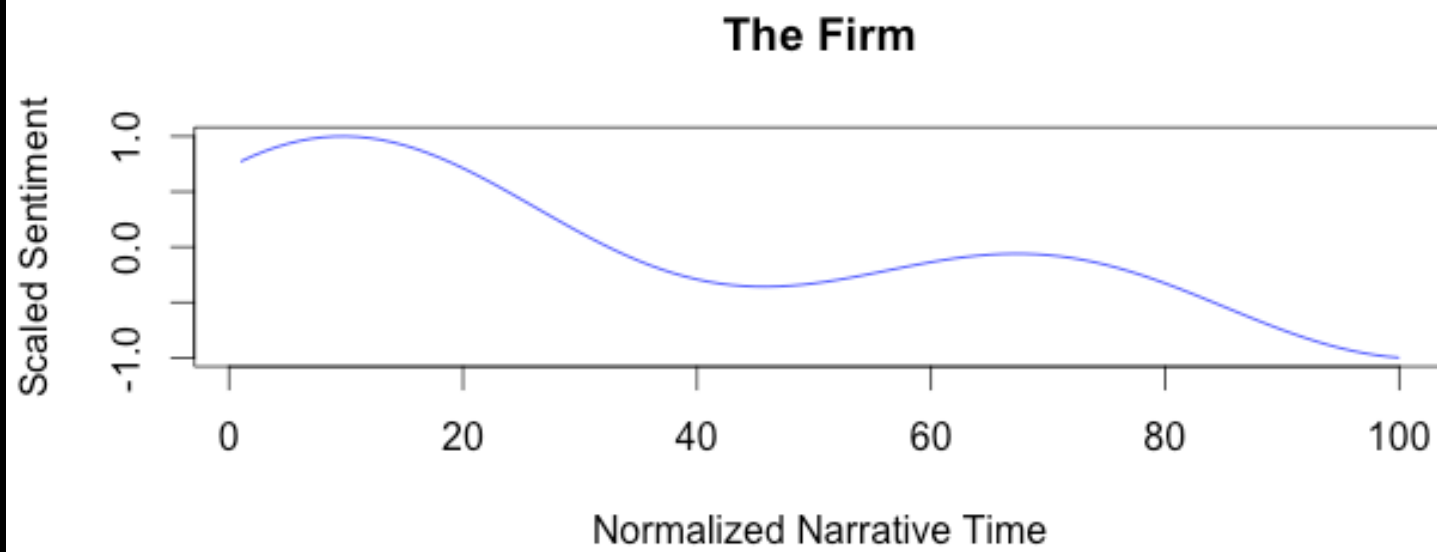
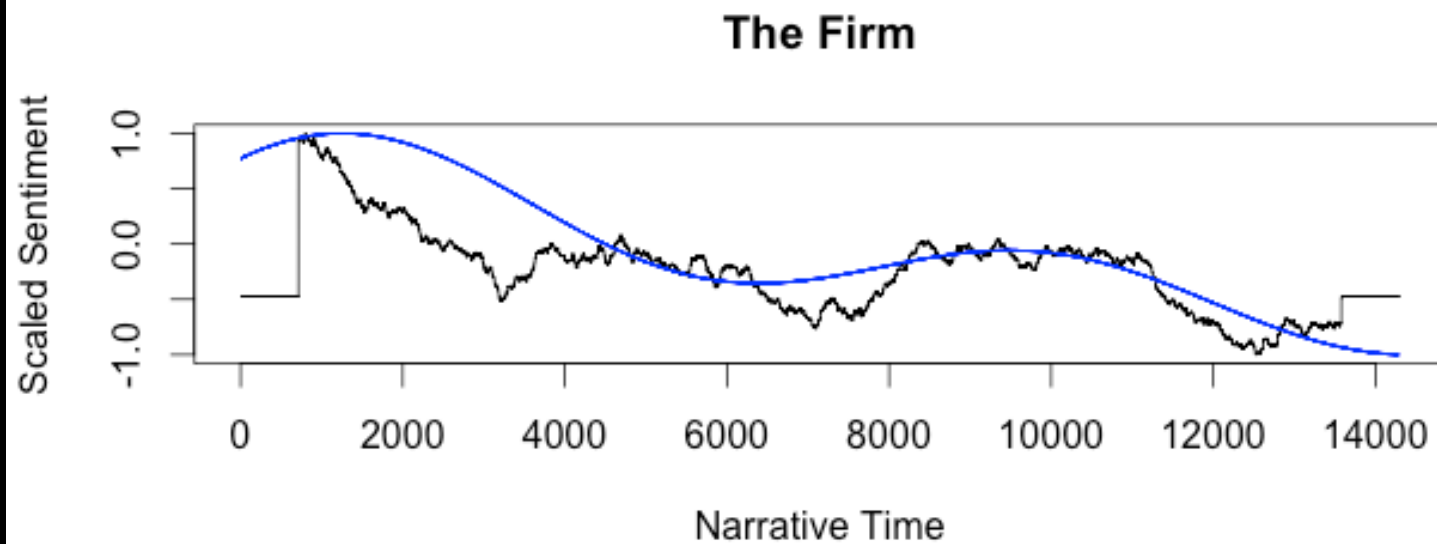
(using DCT suggested by Brad Riddle)

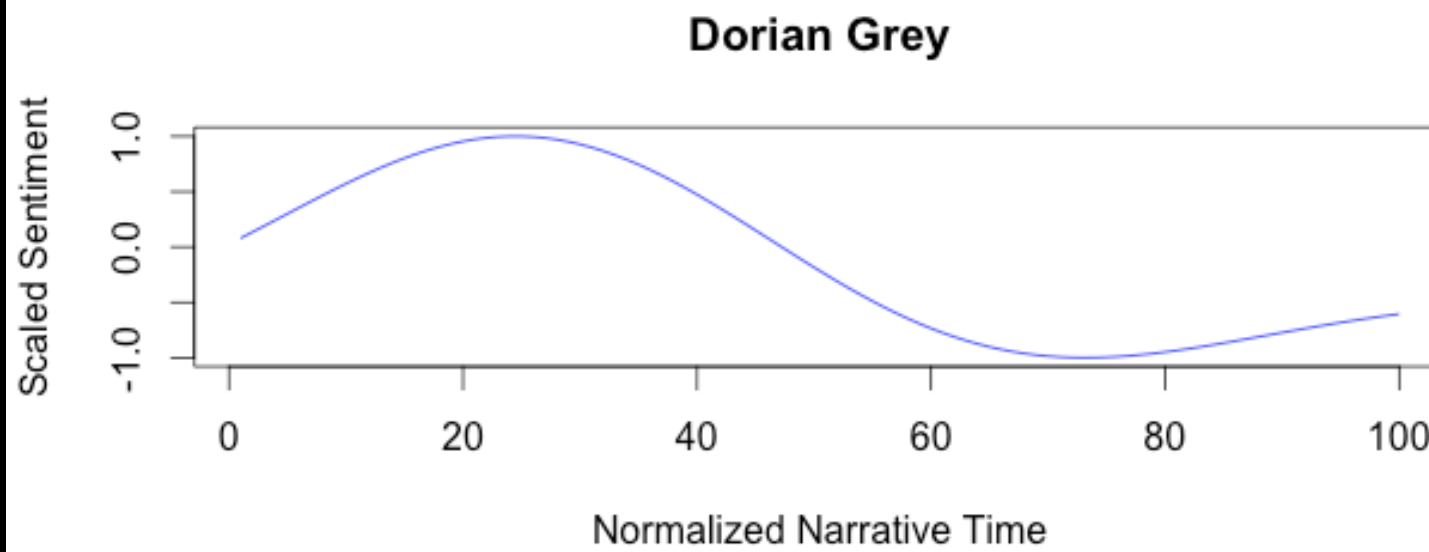
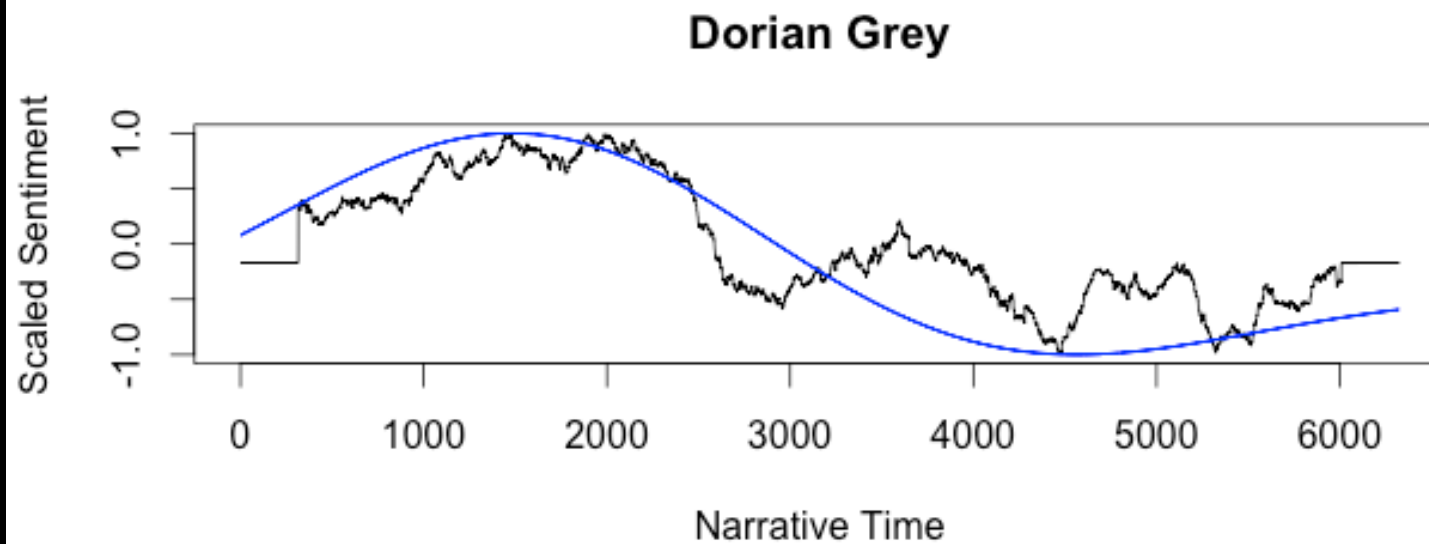




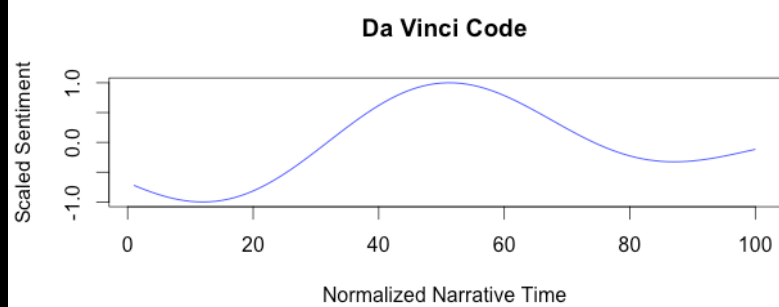
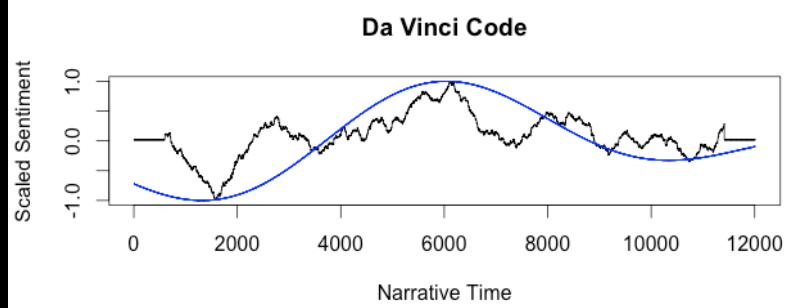
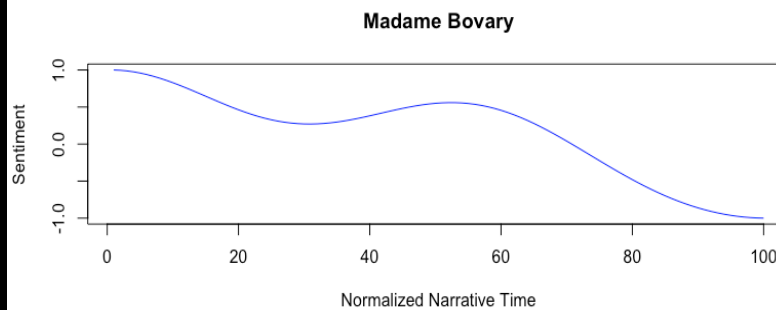
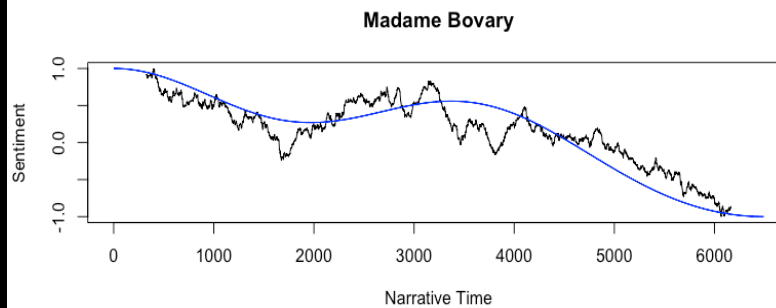
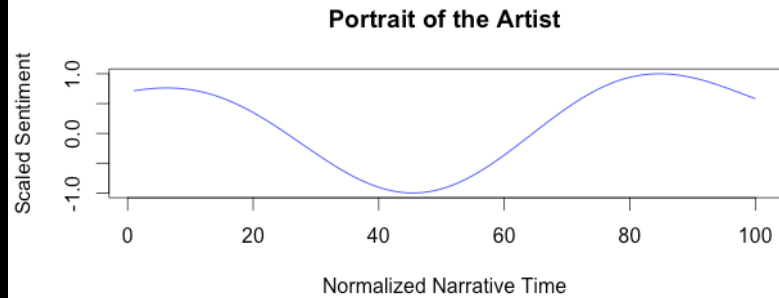
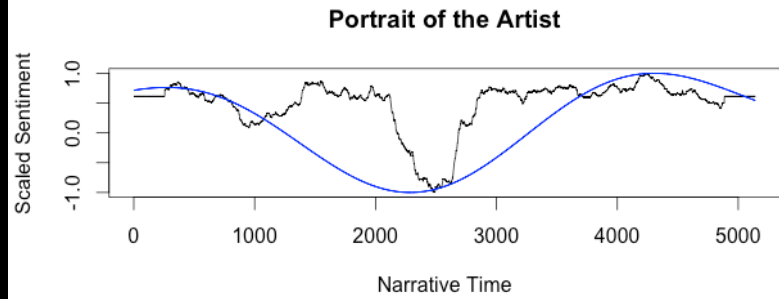
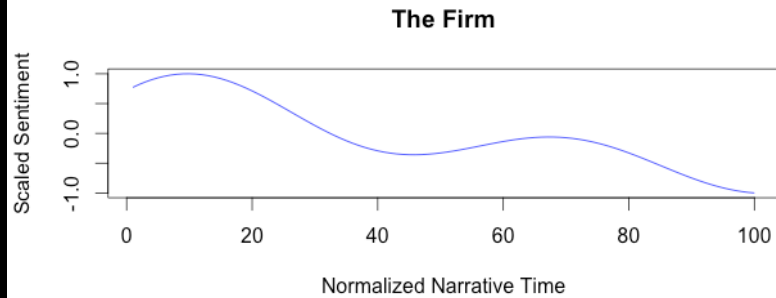
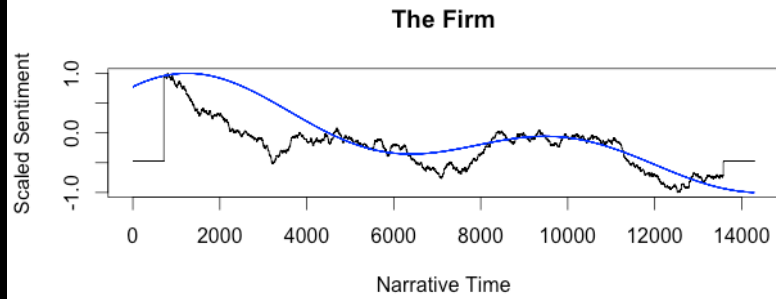




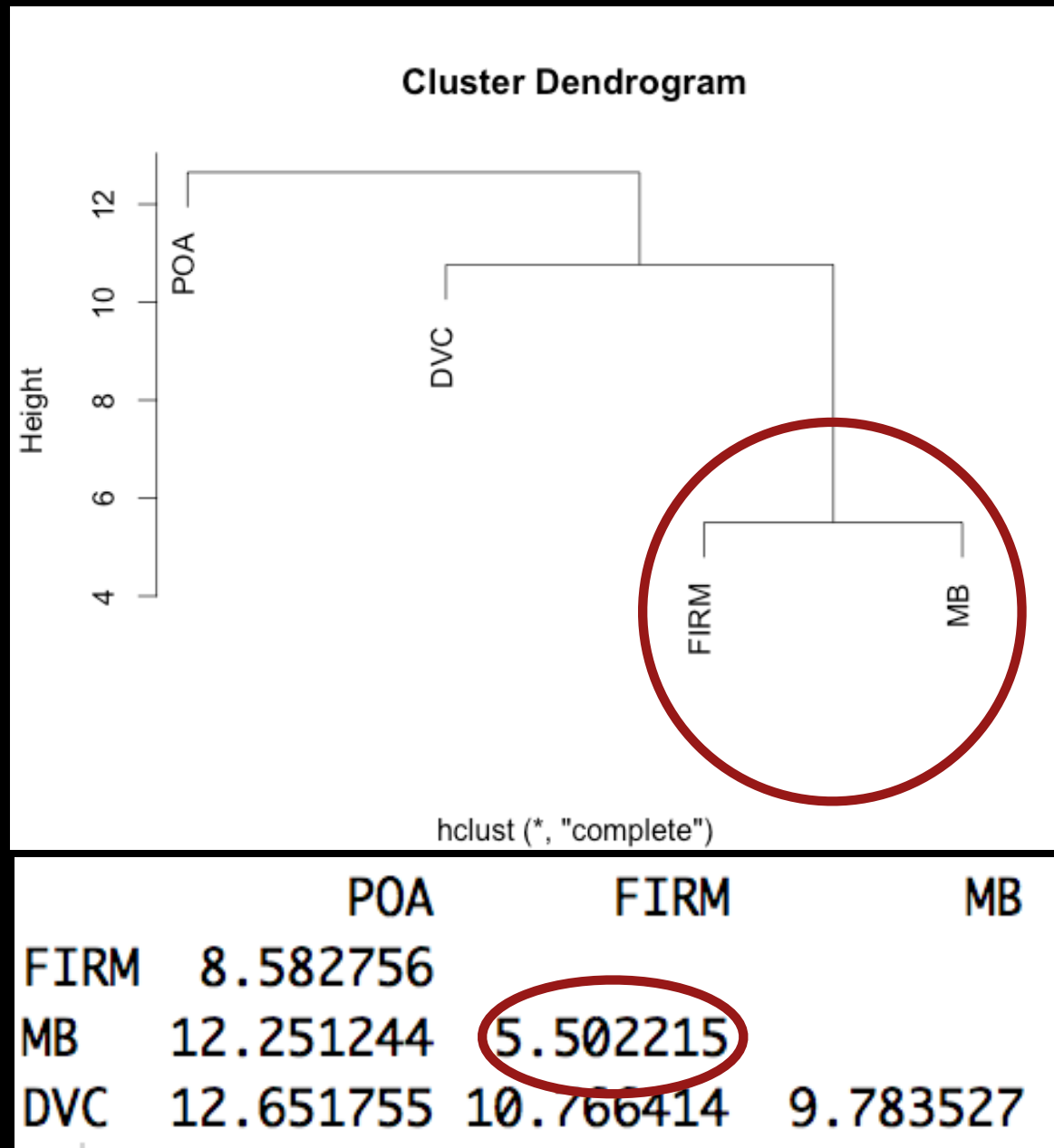




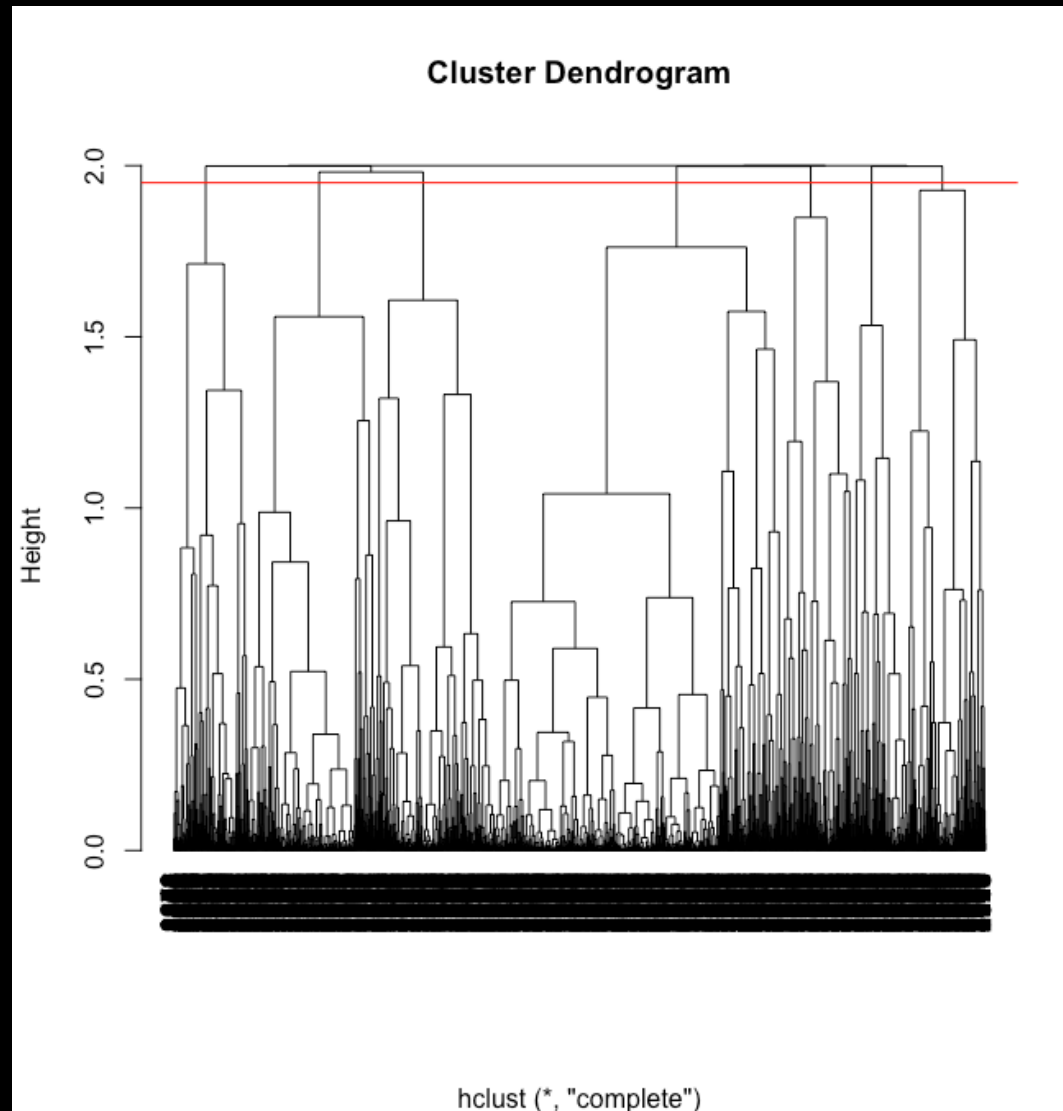
Pick the two most similar



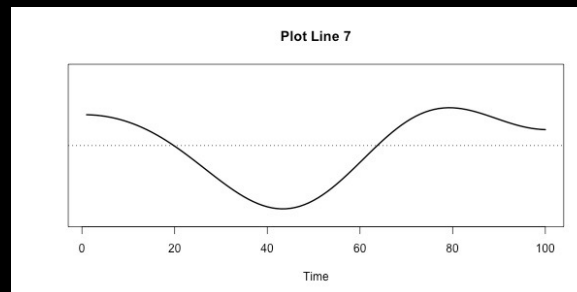
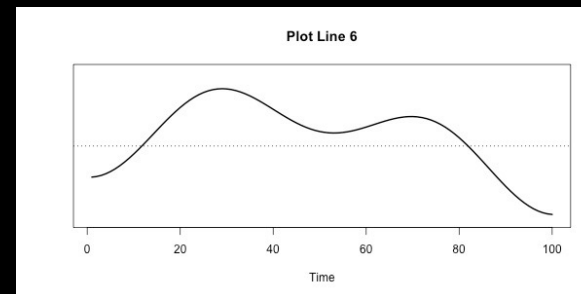
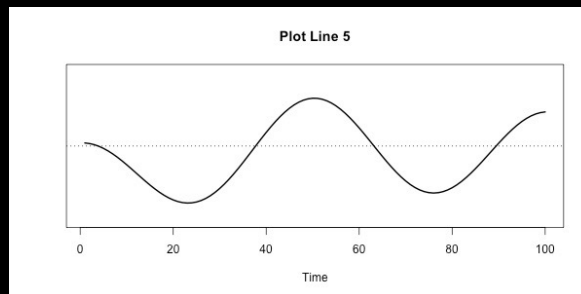
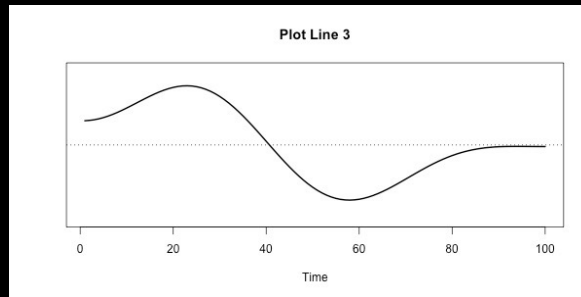
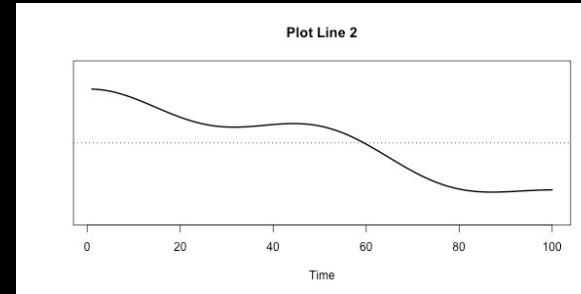
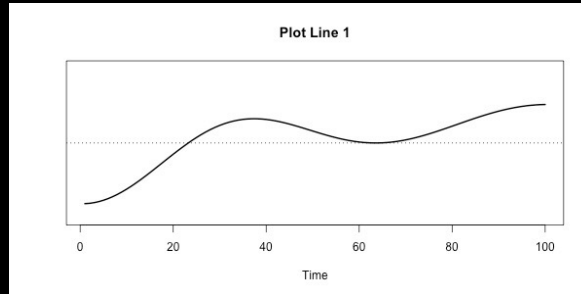
Euclidean Distance



“Ideal” number of clusters = 7



Seven (average) Plot Shapes



Thank you