# BANKING ON SENTIMENT

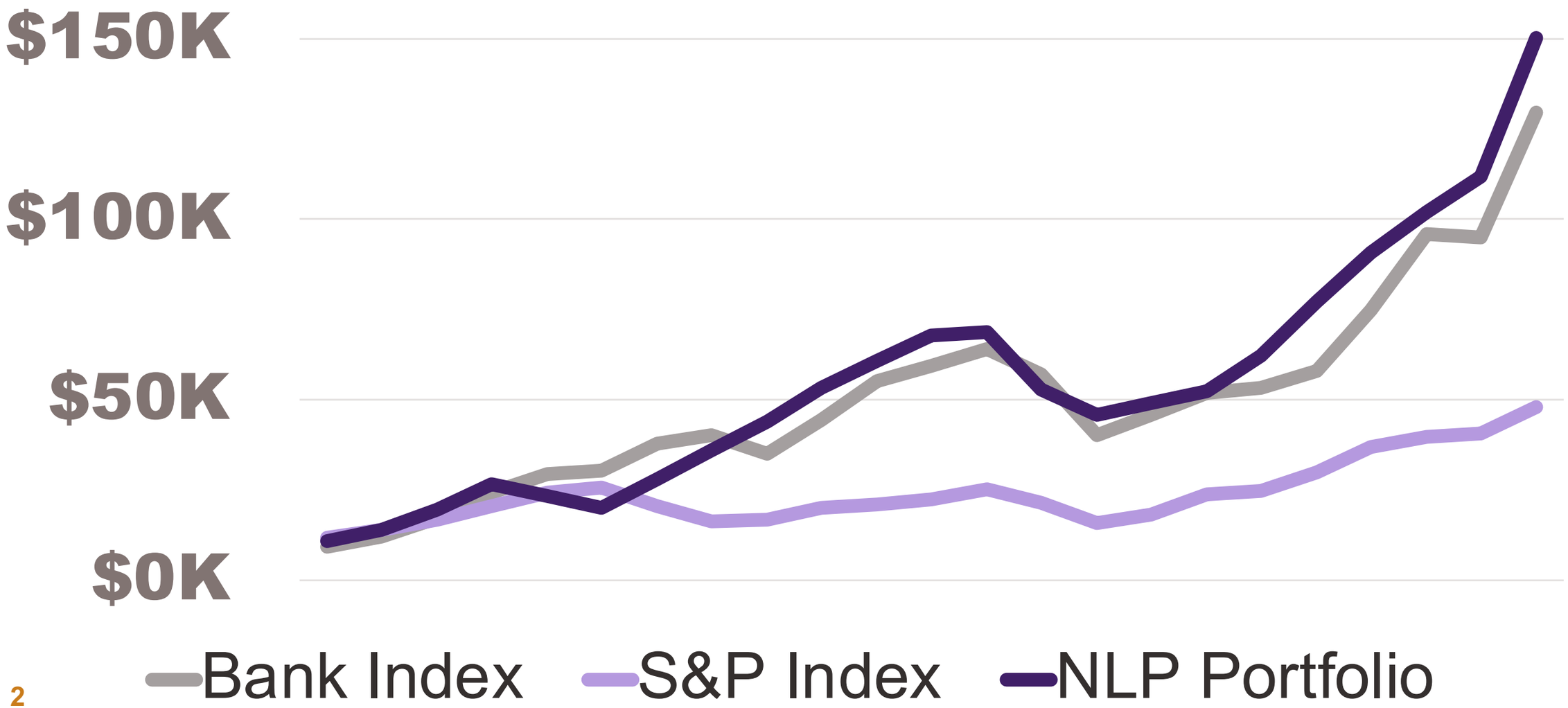
Sandra J.H. Rolnicki, PhD

How NLP* Selected a Bank Stock Portfolio That Outperformed a Bank Stock Index

* Natural Language Processing

2019 R/ Finance Conference

# NLP OUTPERFORMED INDICES



# HOW DID I DO IT?

# BY ANALYZING A LOT OF WORDS

11,000

Annual Reports from Banks

41,000

Average Word Count

| STEPS | R PACKAGES |
| --- | --- |
| Organize | `googleCloudStorageR` |
| Cleanse | `tm` |
| Analyze | `syuzhet` |
| Build | `portfolio` |
| Study | `EventStudy` |

# ORGANIZE

- Obtain annual reports
- Store in Google Cloud

```
library(googleCloudStora
geR)

gcs_upload(object,
  name=”filename",
  bucket = gcs_bucket,
  predefinedAcl =
  "bucketOwnerFullControl
  ")

gcs_get_object(”file",
  bucket = gcs10k_bucket)
```

# CLEANSE

- Remove unwanted characters
- Prep for text analysis

```
library(tm)

str_replace_all(text,
  "[\r\n\t]" , " ")

str_replace_all(text,
  '[\\"*]', "")

removeNumbers(text)

removePunctuation(text)

stripWhitespace(text)
```

# ANALYZE

- Eight sentiment categories*
- Sum for emotional valence (EV)

```
library(syuzhet)

## Array of 10
sentiment <-
  get_nrc_sentiment(text)

## Emotional valence
## Sum first 8 values
emo.val <-
  rowSums(sentiment
  [,1:8])
```

* Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust

“We’re out-performing last year by large margins but we’re concerned about rates.”

**HIGH EV**

“Our results are consistent with our annual projections.”

**LOW EV**

# BUILD

- Stock portfolios
  - Low EV
  - High EV
- Rebalance every July 1

```
library(portfolio)

low.EV.portfolio <-
  new("portfolioBasic",
  instant = i,
  id.var = "cusip",
  in.var = "sentiment",
  type = "equal",
  ret.var = "return",
  data = data)
```

# STUDY

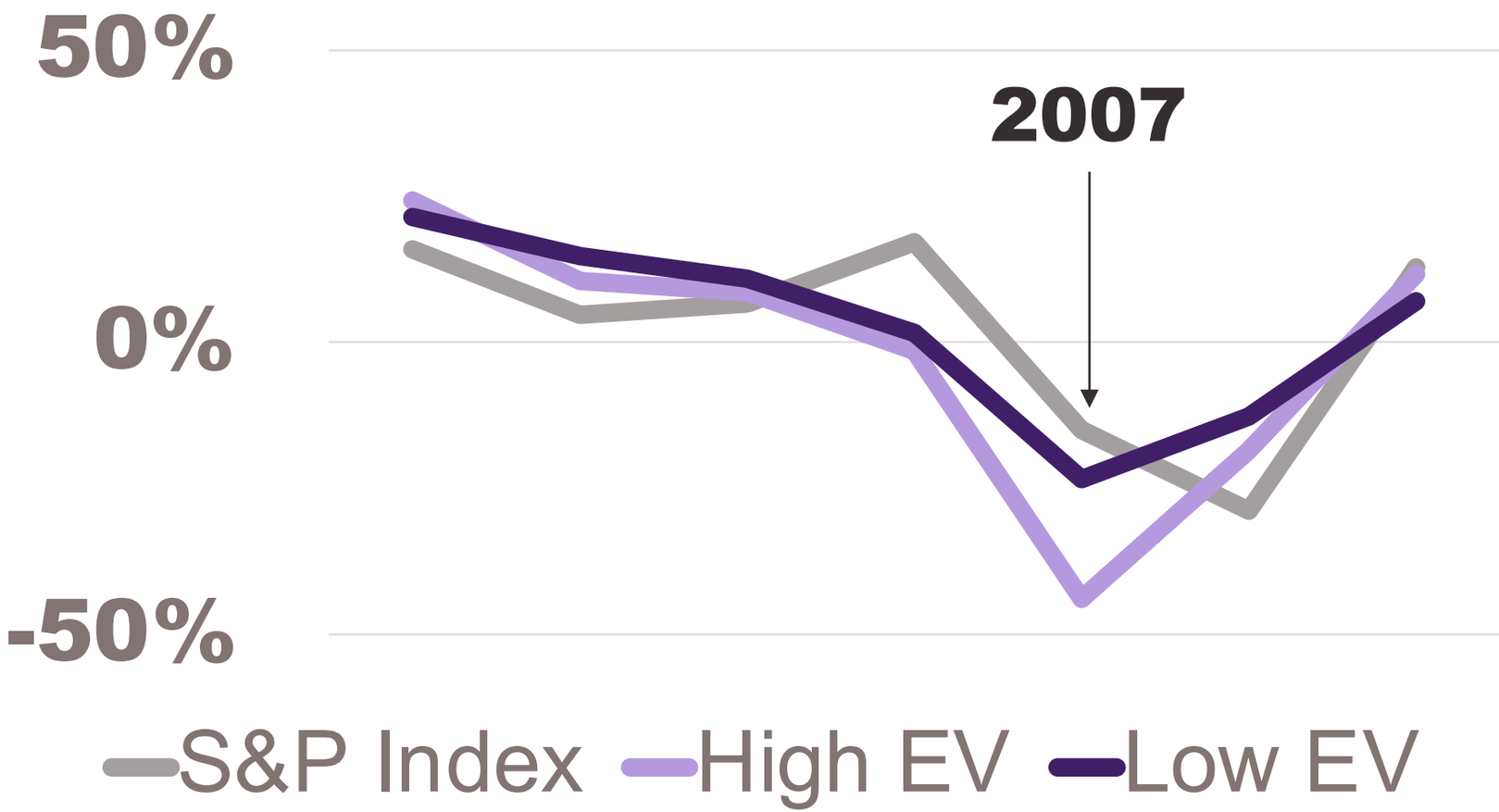
- Perform event study
- Annual report release = Day 0
- 90-day window

```
library(EventStudy)

eventstudy(firm.returns
  = returns,
  event.list = list,
  event.window = 90,
  type = "marketModel",
  to.remap = TRUE,
  remap = "cumsum",
  model.args =
  market.returns))
```

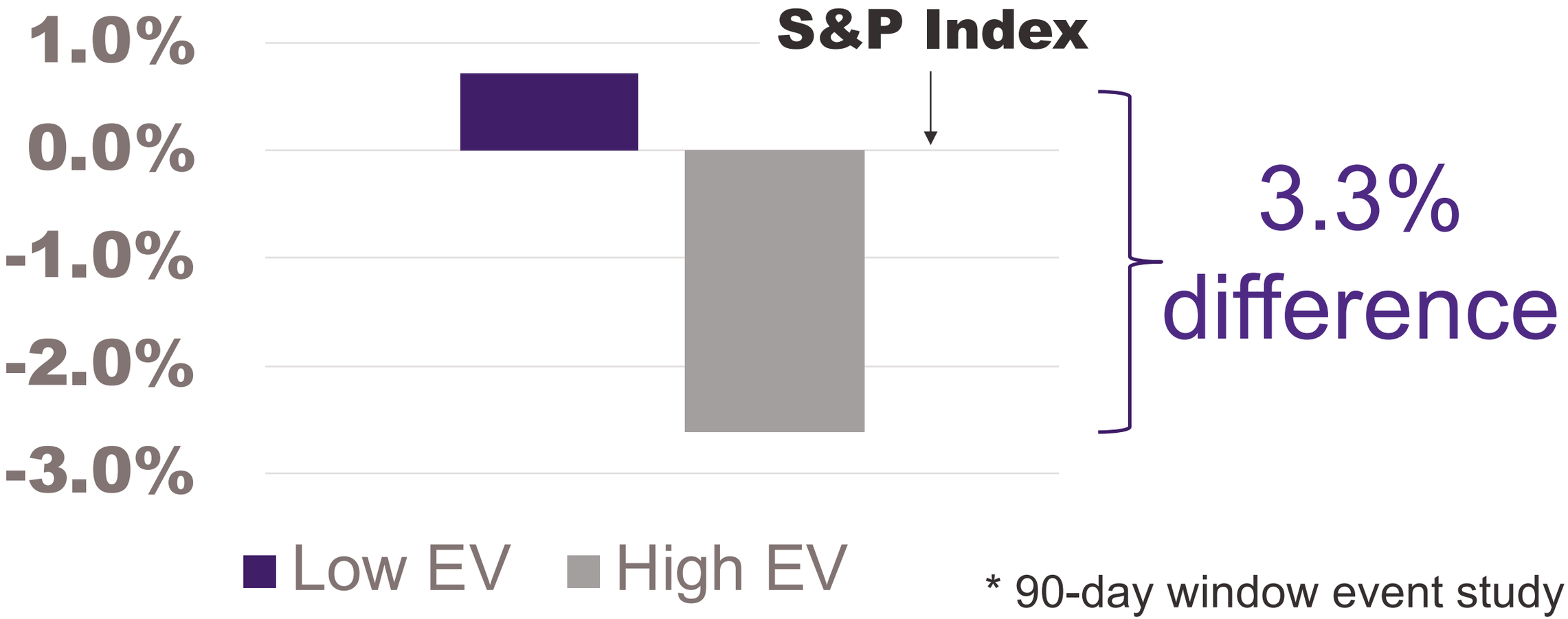
# RESULTS

# CRISIS ERA RETURNS




Leading indicator?

# ABNORMAL RETURNS*




* 90-day window event study

# CONCLUSIONS

- Utility of R packages
- Role of sentiment in annual reports
- Look for the story in your data

# QUESTIONS?

*The author is an adjunct faculty member at Northwestern University and an employee of the Federal Reserve Bank of Chicago. The opinions expressed are her own, and are not formal opinions of, nor binding on, the Federal Reserve Bank of Chicago or the Board of Governors of the Federal Reserve System.*

*All data was obtained from public sources or subscription-based services purchased by Illinois Institute of Technology's Stuart School of Business. This work does not contain confidential supervisory information in detail or in aggregate.*

# CONNECT

https://www.linkedin.com/in/sjhrolnicki/

https://orcid.org/0000-0002-4849-6880

https://github.com/Sandra-R-PhD

# Bibliography

Edmondson, M. (2017). googleCloudStorageR: Interface with Google Cloud Storage API. R package version 0.4.0, URL: https://CRAN.R-project.org/package=googleCloudStorageR.

Enos, J. & Kane, D. (2015). portfolio: Classes for analysing and implementing equity portfolios. R package version 0.4-7, URL: https://CRAN.R-project.org/package=portfolio.

# Bibliography

Feinerer, I. & Hornik, K. (2018). tm: Text Mining Package. R package version 0.7-6, URL: https://CRAN.R-project.org/package=tm.

Jockers, M. (2017). syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. R package version 1.0.4, URL: https://CRAN.R-project.org/package=syuzhet.

# Bibliography

Mueller, S. (2019). EventStudy: Event Study Analysis. R package version 0.36, URL: https://CRAN.R-project.org/package=EventStudy.