Matthew Cocci's Potent Productivity

Micah Smith

July 24, 2015

1. Introduction
   - Historical productivity
   - Motivation of empirical study
   - Description of data
2. Pretty tables/pictures
   - Summary statistics
   - Graphical trends
3. Model
4. Conclusions

Cocci is very productive

- This is pretty obvious to most of us.

Cocci is very productive

- This is pretty obvious to most of us.
- How does he do it?

Cocci is very productive

- This is pretty obvious to most of us.
- How does he do it?
- No one really knows.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.
- Conduct inference on causes (and *effects*) of productivity.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.
- Conduct inference on causes (and *effects*) of productivity.
  - Cocci is a big fan of the podcast Serial.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.
- Conduct inference on causes (and *effects*) of productivity.
  - Cocci is a big fan of the podcast Serial.
  - Presumably, his productivity decreased on days in which he was listening to a new Serial episode.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.
- Conduct inference on causes (and *effects*) of productivity.
  - Cocci is a big fan of the podcast Serial.
  - Presumably, his productivity decreased on days in which he was listening to a new Serial episode.
  - If we could analyze his productivity over time, we could determine on which days new episodes of Serial were released.

Motivation of empirical study

- Collect data to document productivity for future generations of RAs.
- Conduct inference on causes (and *effects*) of productivity.
  - Cocci is a big fan of the podcast Serial.
  - Presumably, his productivity decreased on days in which he was listening to a new Serial episode.
  - If we could analyze his productivity over time, we could determine on which days new episodes of Serial were released.
- Embarrass him in front of all his loved ones.

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.
- I collect a novel dataset of hourly observations of the number of panes open:

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.
- I collect a novel dataset of hourly observations of the number of panes open:
  - Timestamped at 9am through 5pm, seven days a week.

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.
- I collect a novel dataset of hourly observations of the number of panes open:
  - Timestamped at 9am through 5pm, seven days a week.
  - Data collected from August 11, 2014 to yesterday (July 21, 2015).

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.
- I collect a novel dataset of hourly observations of the number of panes open:
  - Timestamped at 9am through 5pm, seven days a week.
  - Data collected from August 11, 2014 to yesterday (July 21, 2015).
  - 3100 hourly observations over 345 days in total.

Description of data

- Cocci uses the `tmux` utility on the RAN to manage large collections of "panes", or pseudo terminal sessions.
- Make observations of the number of panes that Cocci has open at a given point to get a sense of productivity over time.
- I collect a novel dataset of hourly observations of the number of panes open:
  - Timestamped at 9am through 5pm, seven days a week.
  - Data collected from August 11, 2014 to yesterday (July 21, 2015).
  - 3100 hourly observations over 345 days in total.
  - Don't even worry about how I did this.

Summary statistics

Table 1: Description of hourly data

	panes	absd1panes	sqd1panes
count	2181.00	2180.00	2180.00
mean	23.44	1.06	5.52
std	12.66	2.09	30.54
min	0.00	0.00	0.00
25%	15.00	0.00	0.00
50%	21.00	0.00	0.00
75%	30.00	1.00	1.00
max	73.00	26.00	676.00

Hours excluded from sample are non-workdays or days on which the RAN was reset.

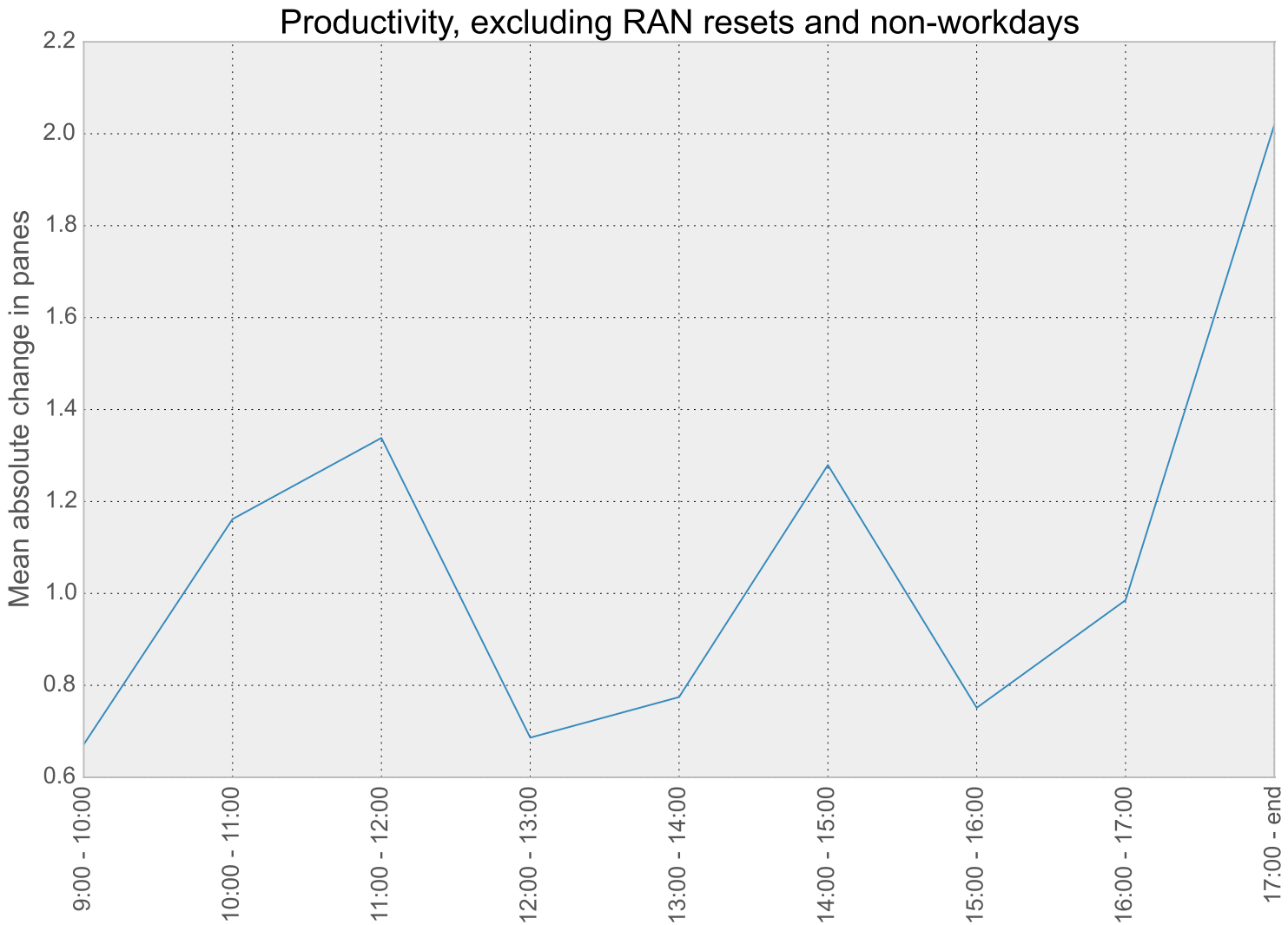
Summary statistics

Table 2: Description of daily (collapsed) data

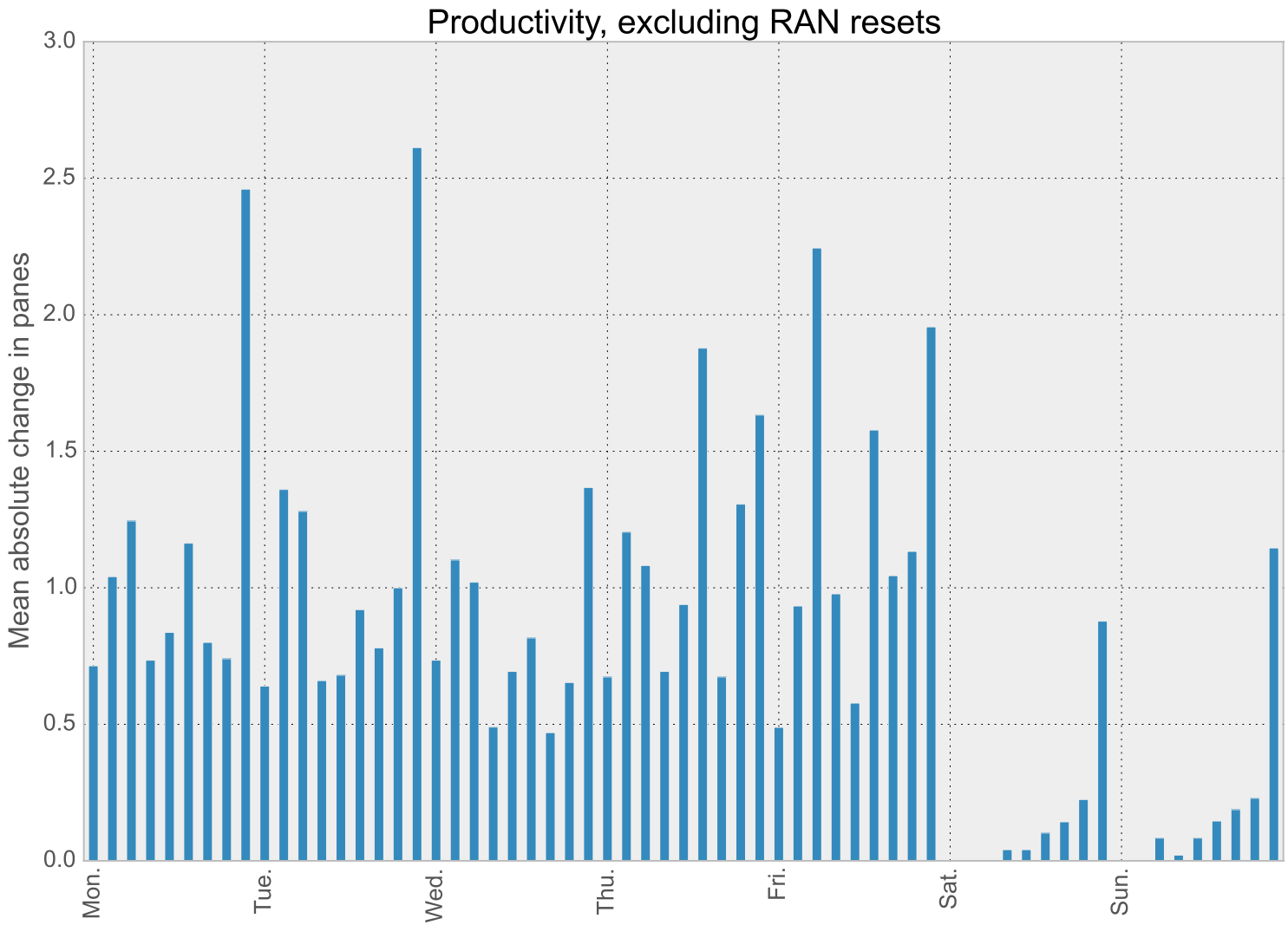
	has_episode_today	absd1panes	sqd1panes
count	243	243.00	243.00
mean	0.04	9.55	49.49
std	0.20	7.94	95.27
min	False	0.00	0.00
25%	0.00	4.00	6.00
50%	0.00	8.00	18.00
75%	0.00	13.00	54.00
max	True	46.00	756.00

Days excluded from sample are non-workdays or days on which the RAN was reset. `has_episode_today` is marked if an episode of Serial was released on that day.

Hourly trend



Weekly trend



Model: Description

- Use absolute change in panes as a rough proxy for hourly productivity, and aggregated absolute hourly change in panes as a rougher proxy for daily productivity.

Model: Description

- Use absolute change in panes as a rough proxy for hourly productivity, and aggregated absolute hourly change in panes as a rougher proxy for daily productivity.
- Include squared term to capture all of the super duper non-linear effects we expect.

Model: Description

- Use absolute change in panes as a rough proxy for hourly productivity, and aggregated absolute hourly change in panes as a rougher proxy for daily productivity.
- Include squared term to capture all of the super duper non-linear effects we expect.
- Control for day of the week as well as Brent Crude spot prices.
- Train a boosted decision trees classifier, because why not, and because Cocci deserves only the best.

Model: Description

- Use absolute change in panes as a rough proxy for hourly productivity, and aggregated absolute hourly change in panes as a rougher proxy for daily productivity.
- Include squared term to capture all of the super duper non-linear effects we expect.
- Control for day of the week as well as Brent Crude spot prices.
- Train a boosted decision trees classifier, because why not, and because Cocci deserves only the best.
- Train on data from August 11, 2014 to March 31, 2015; test on data from April 1 to yesterday.

Model: Results

Table 3: Days with potential Serial episode releases

	pred
date	
2015-04-02	0.54
2015-05-12	0.90
2015-06-18	0.62
2015-07-02	0.54
2015-07-16	0.54

Last *known* episode of Serial released on December 18, 2014.

Conclusions

- Based on Cocci's productivity trends, an episode of Serial was released with 90% certainty on May 12. It's just that no one but him has noticed yet.

Conclusions

- Based on Cocci's productivity trends, an episode of Serial was released with 90% certainty on May 12. It's just that no one but him has noticed yet.
- Good luck at Princeton!

Conclusions

- Based on Cocci's productivity trends, an episode of Serial was released with 90% certainty on May 12. It's just that no one but him has noticed yet.
- Good luck at Princeton!
- Do Something Good Everyday!