

Open Data Source Analysis

Daniel “whopis” Burroughs

daniel.burroughs@gmail.com

About me...

- Research Professor
 - University of Central Florida
 - College of Engineering
 - College of Health and Public Affairs
- Associate Technology Director
 - Center for Law Enforcement Technology, Training & Research (LETTR)

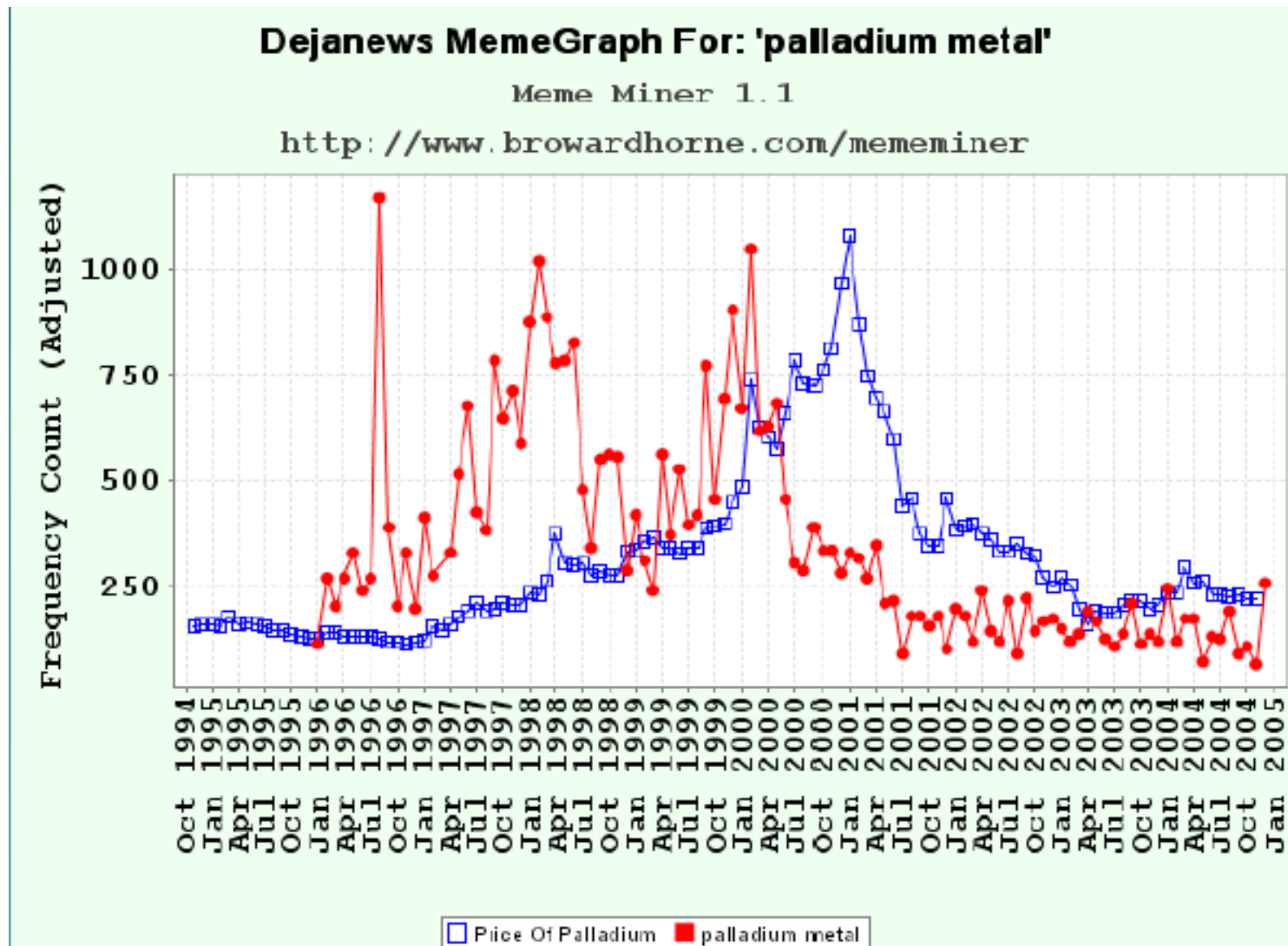
Related Interests

- Pattern recognition in security applications
 - Network attack detection and classification
 - Bayesian Multiple Hypothesis Tracking
 - Detection avoidance
- Sensor fusion
 - Position tracking & estimation
 - Autonomous navigation
 - Situational awareness

My Inspiration

- Broward Horne
 - Meme Mining for Fun and Profit
 - DefCon 13
- Detection and identification of saturation curves
 - Which current technologies are about to take off?
 - What should I spend my time learning?

An Example



(blatently stolen from Broward Horne's DC13 talk)

My Need

- When I used to teach...
 - Students would come to me seeking job search advice
 - I would do what I could to keep up with trends
 - Found that craigslist postings were an interesting indicator of what was going on
 - Not necessarily the best source for jobs – but definitely an interesting sensor

(a tiny bit of) Theory

- Jobs are a lagging economic indicator
 - There is a time period between economies changing and the job markets catching up
 - Hiring someone tends to be long term investment
- A lot of craigslist postings are temporary / contract / side jobs
 - Short term investment
 - Shorten the lag

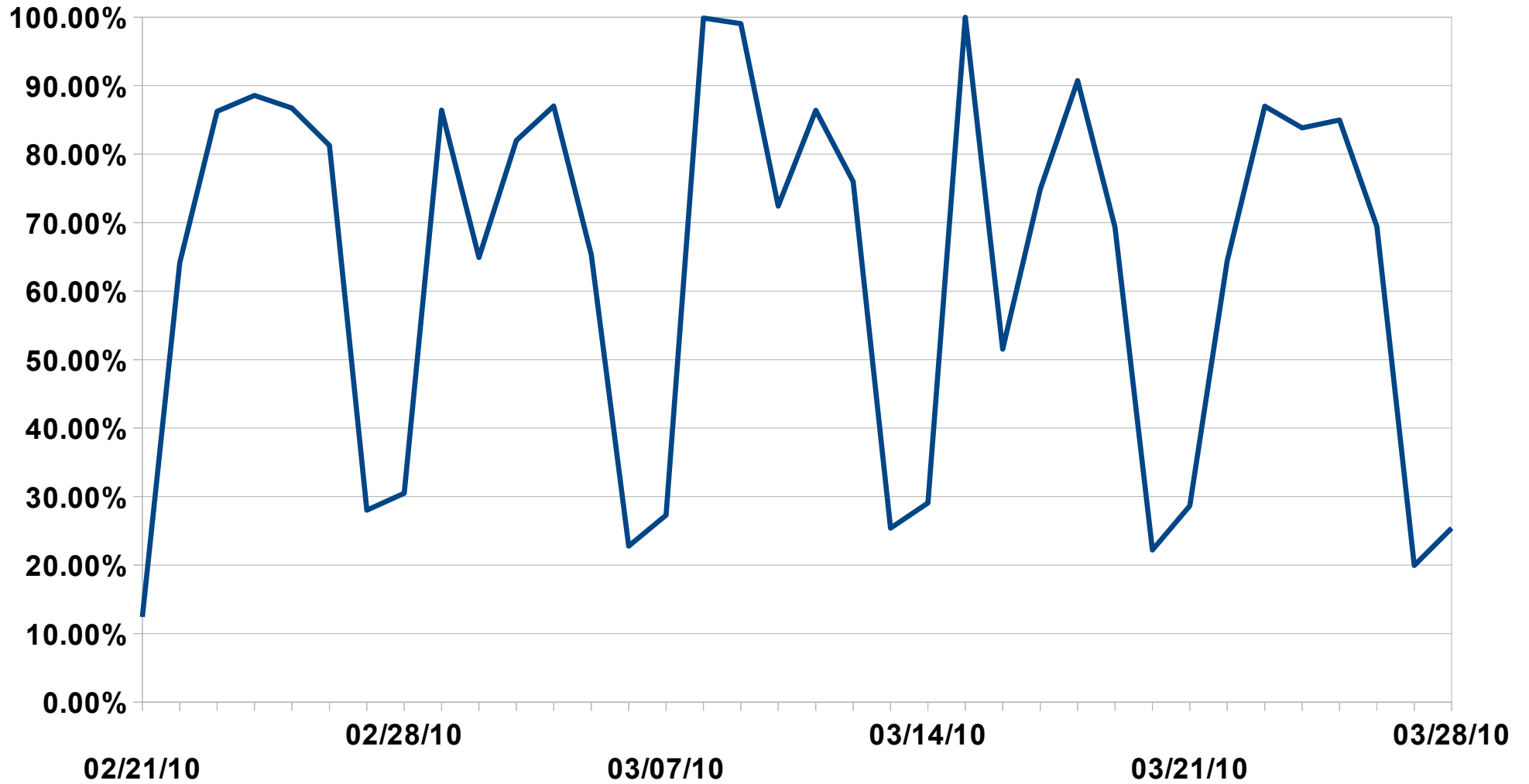
Why Craigslist?

- It's available
- Decent categorization
 - 31 job categories, 400 cities in North America
- More than just job listings
- Significant amount of short-term relevance
 - Both in job listings and other items & services
- Prediction:
 - This will be a normal to leading indicator of change
- This is not the only or best source for data – it is just one example of the many out there

Let's look at some data

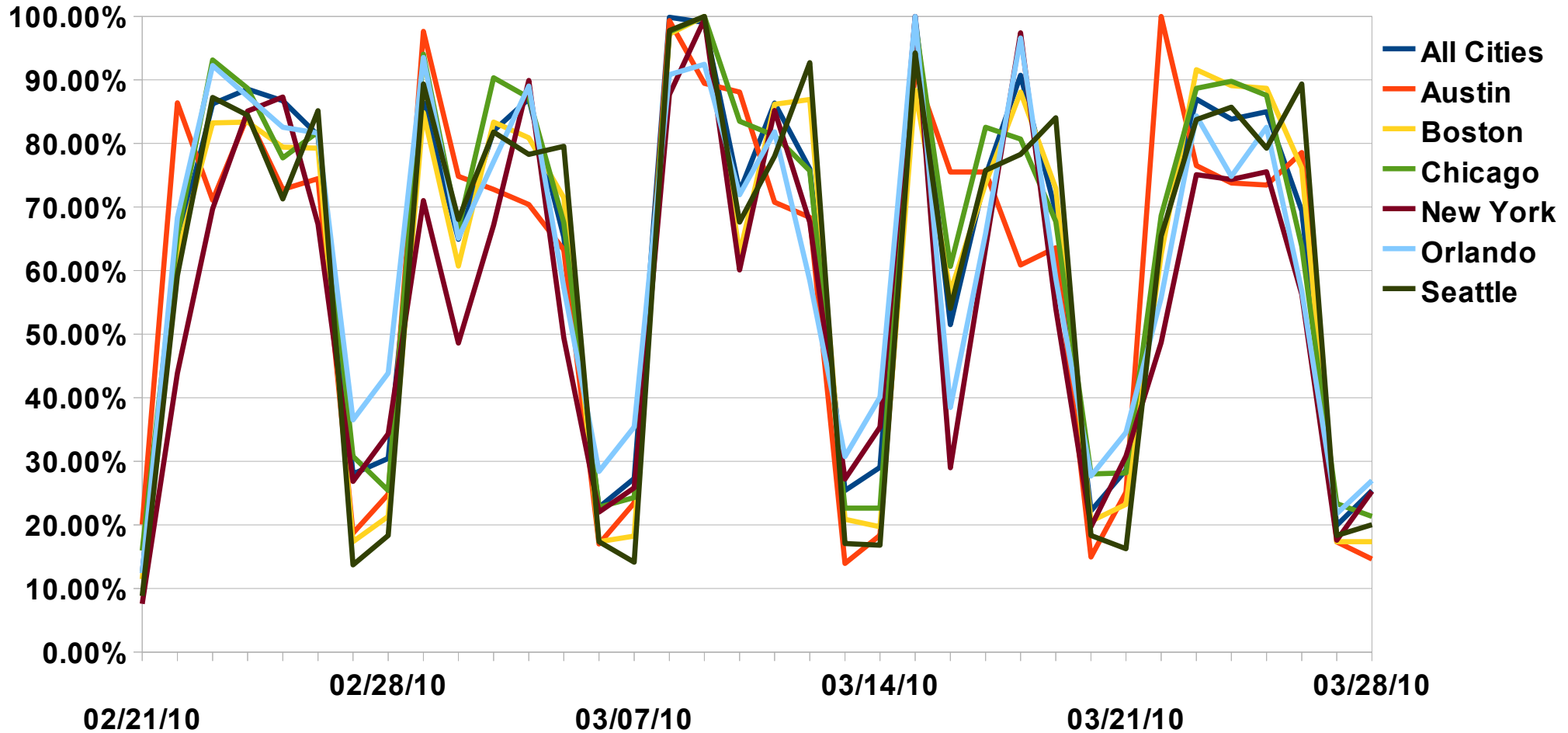
- Initial results
- The easy-to-find stuff
- More questions than answers
 - But that is the fun of it

Job Postings By Date



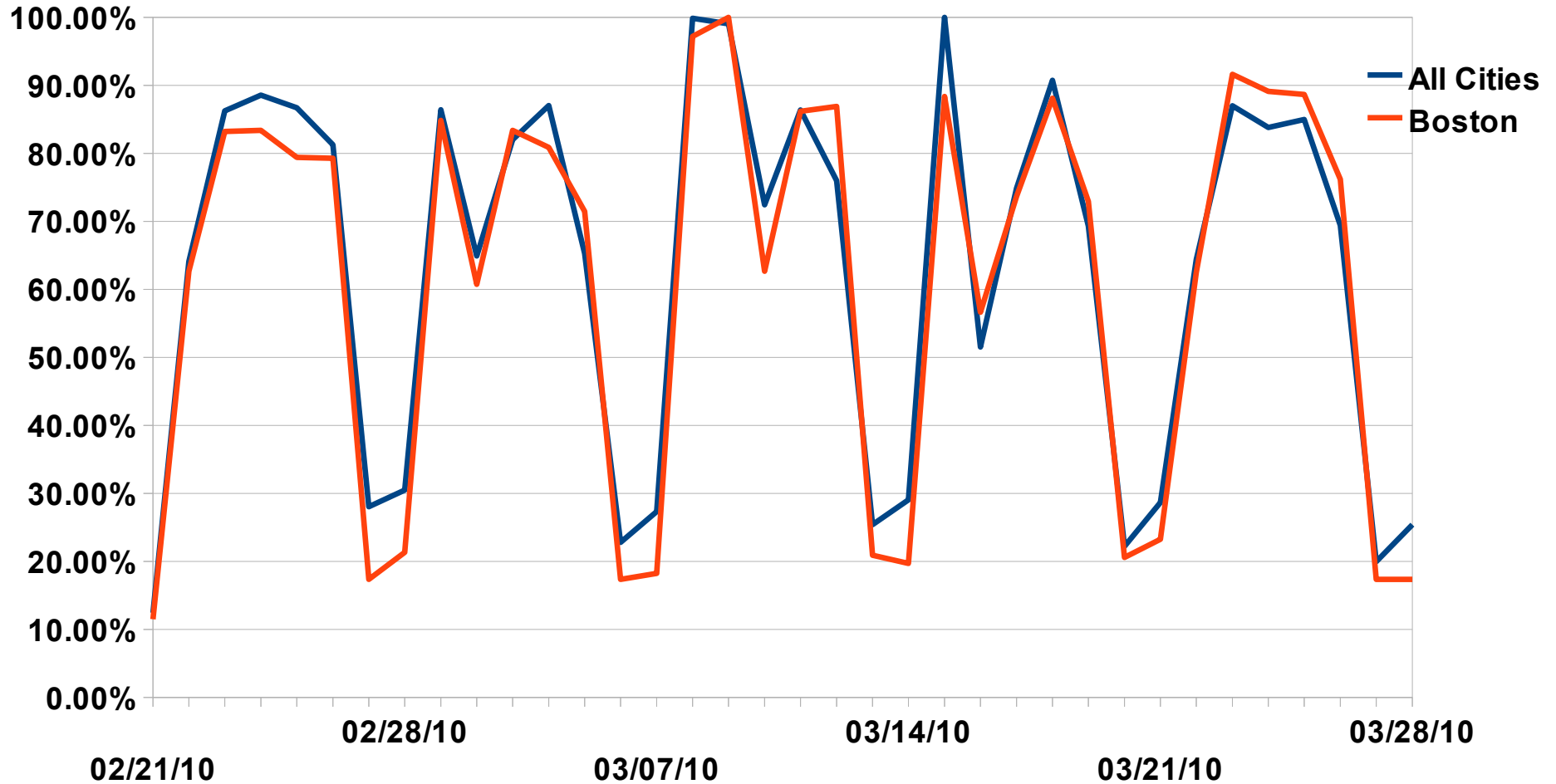
All Categories – All Cities

Job Postings By Date



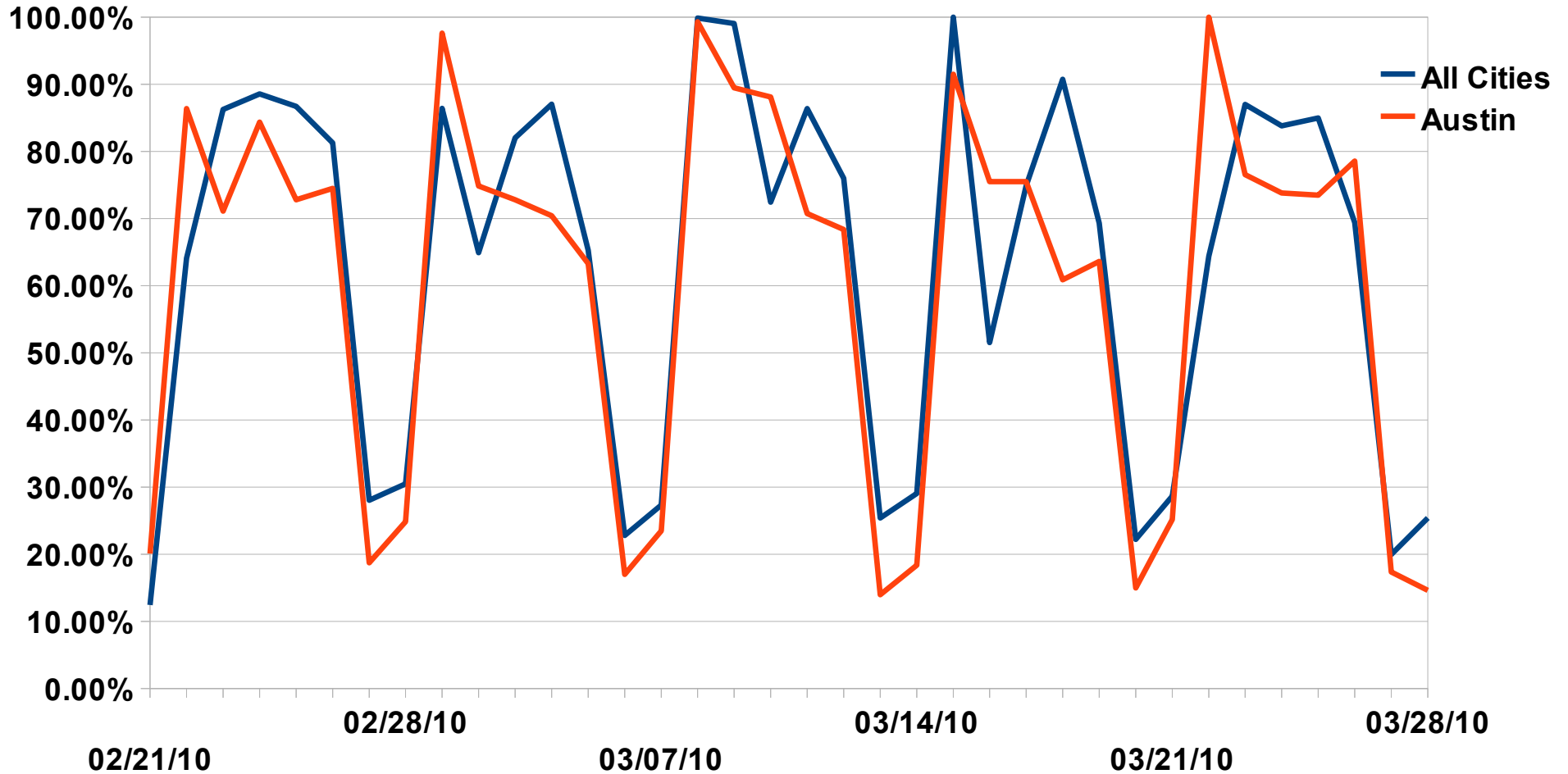
All Categories – Selected Cities

Job Postings By Date



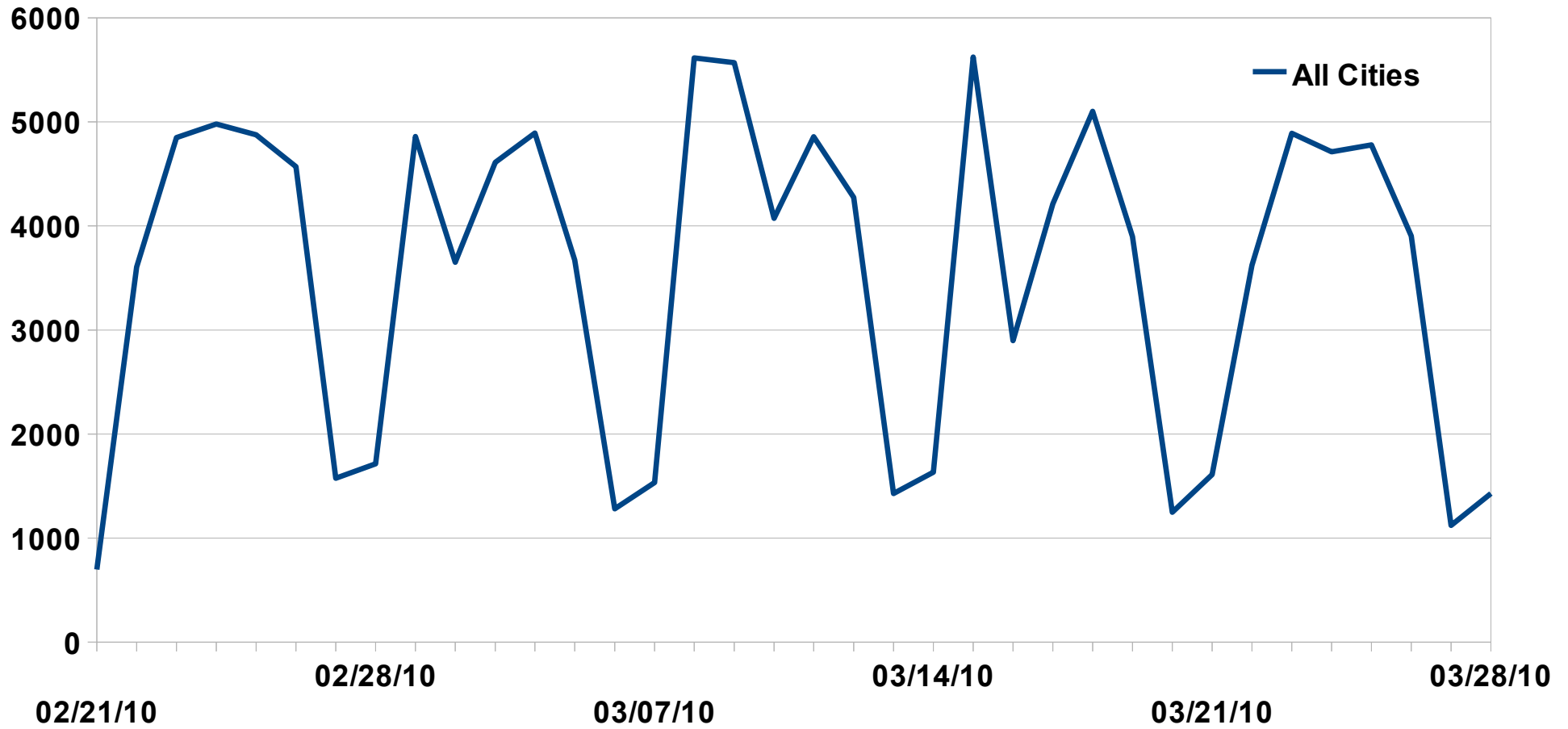
All Categories – Boston

Job Postings By Date



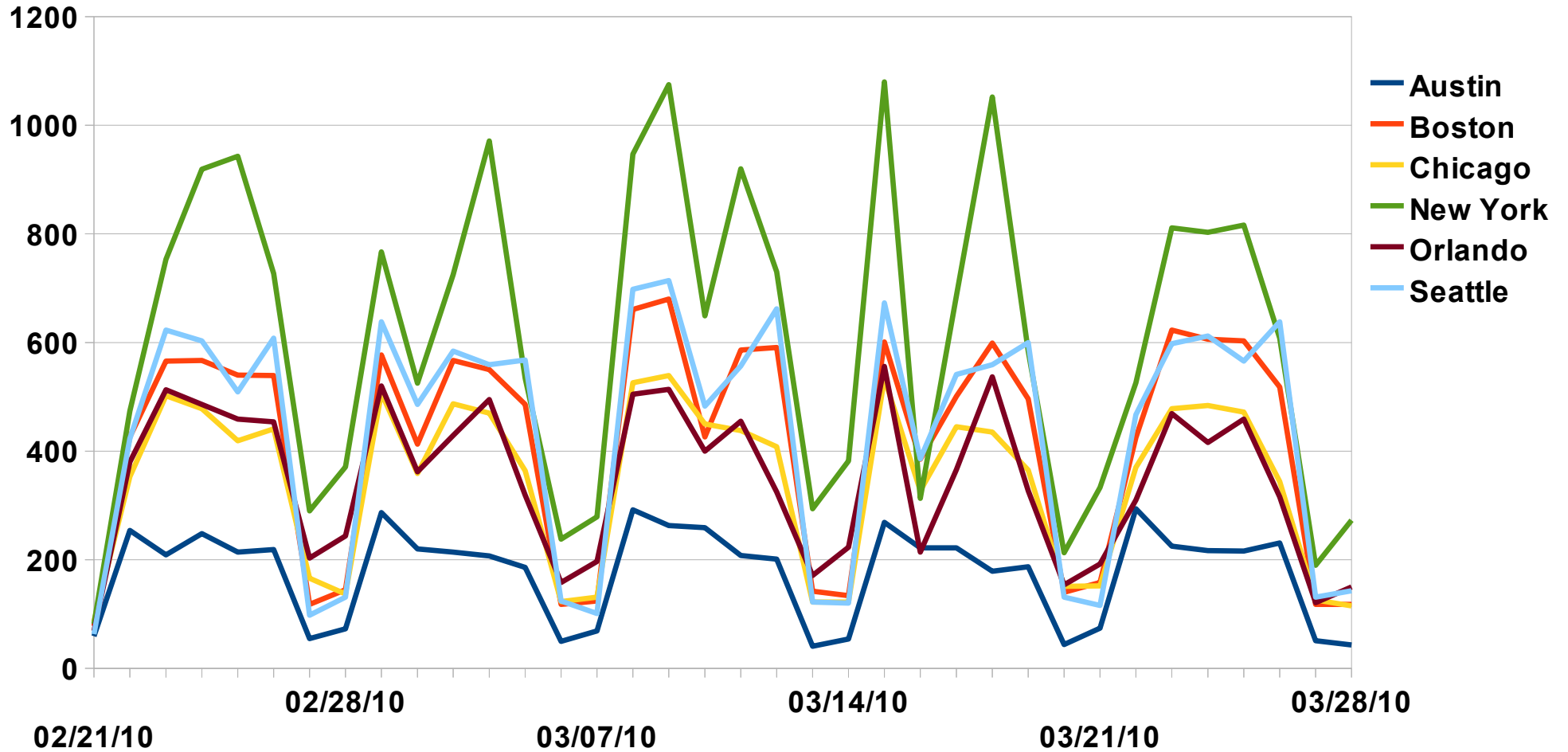
All Categories – Austin

Number of Jobs Posted



All Categories – All Cities

Number of Jobs Posted



All Categories – Selected Cities

Conclusions

- This is in the very early stages
- Clearly there is some significance here
- My goal is to throw this idea out and see what else can be done with it
- There is all sorts of interesting information hidden in the data that is being collected all around us – and huge amounts are openly available – what can we find in it?

Some relevant, some not-so-relevant links and acknowledgments

- www.danielburroughs.com
- www.realmeme.com
- www.hoverflytech.com
- www.gotootie.com
- www.medicaldatasharing.com
- www.lettr.org
- http://www.defcon.org/images/defcon-13/dc13-presentations/DC_13-Horne.pdf