

Statistics 105: Statistics for Engineers

Mark Hansen

e: cocteau@stat.ucla.edu

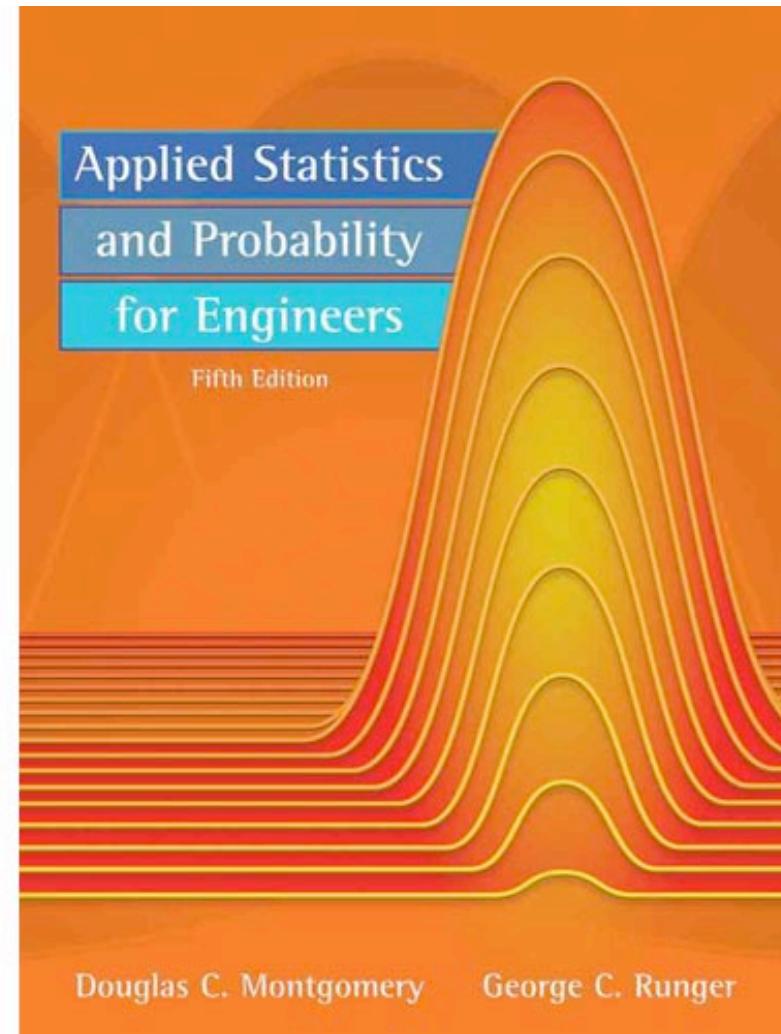
t: @cocteau



TA: Miles Chen



Reader: Ben Greenspan



Chapter 6. Random Sampling and Data Description.

Chapter 7. Point Estimation of Parameters.

Chapter 8. Statistical Intervals for a Single Sample.

Chapter 9. Tests of Hypotheses for a Single Sample.

Chapter 10. Statistical Inference for Two Samples.

Chapter 11. Simple Linear Regression and Correlation.

Chapter 12. Multiple Linear Regression.

Chapter 13. Design and Analysis of Single-Factor Experiments: The Analysis of Variance.

Chapter 14. Design of Experiments with Several Factors.

Chapter 15. Nonparametric Statistics.

Grading

Homework (10 assignments) 50%

Final project 40%

Participation 10%

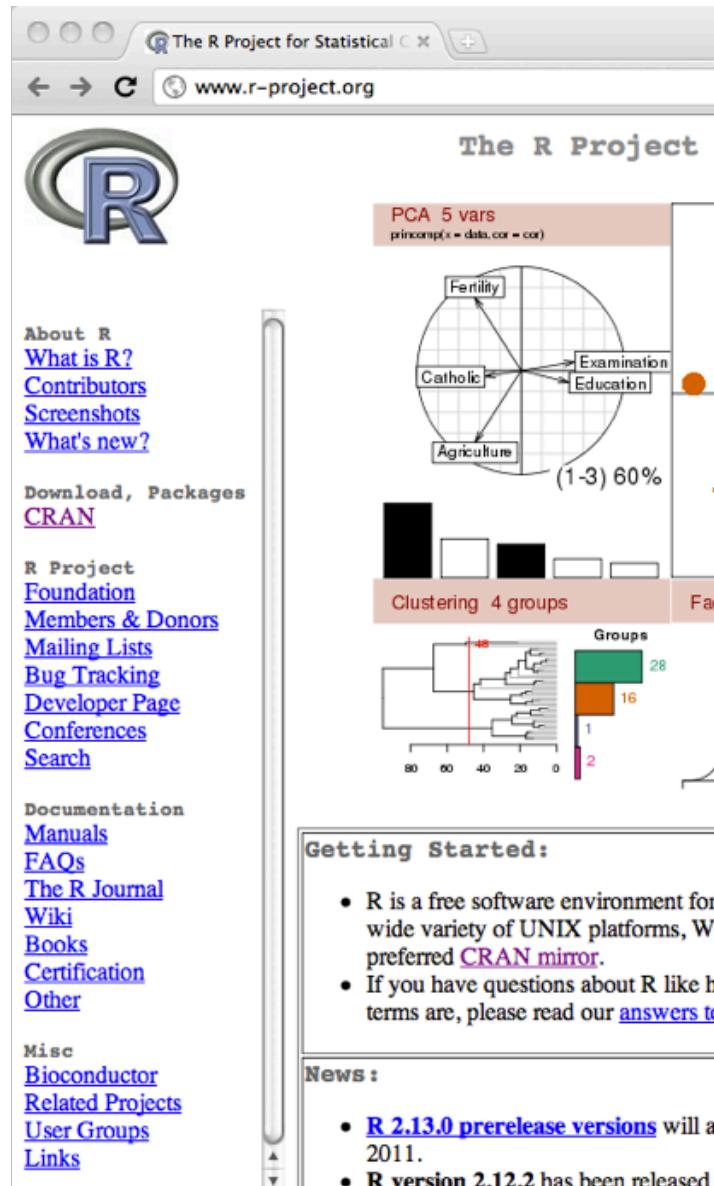
All work is to be submitted via the CCLE course site by the indicated deadline with no exceptions (this sounds harsh, but things get lost otherwise)

You are expected to work on your assignments and your final project alone - We will have more to say about the format of your assignments later, but broadly they will be writing assignments with a code appendix

Computing: The R environment

This quarter, we will perform all of our analysis in the R environment for statistical computing -- We've selected it because

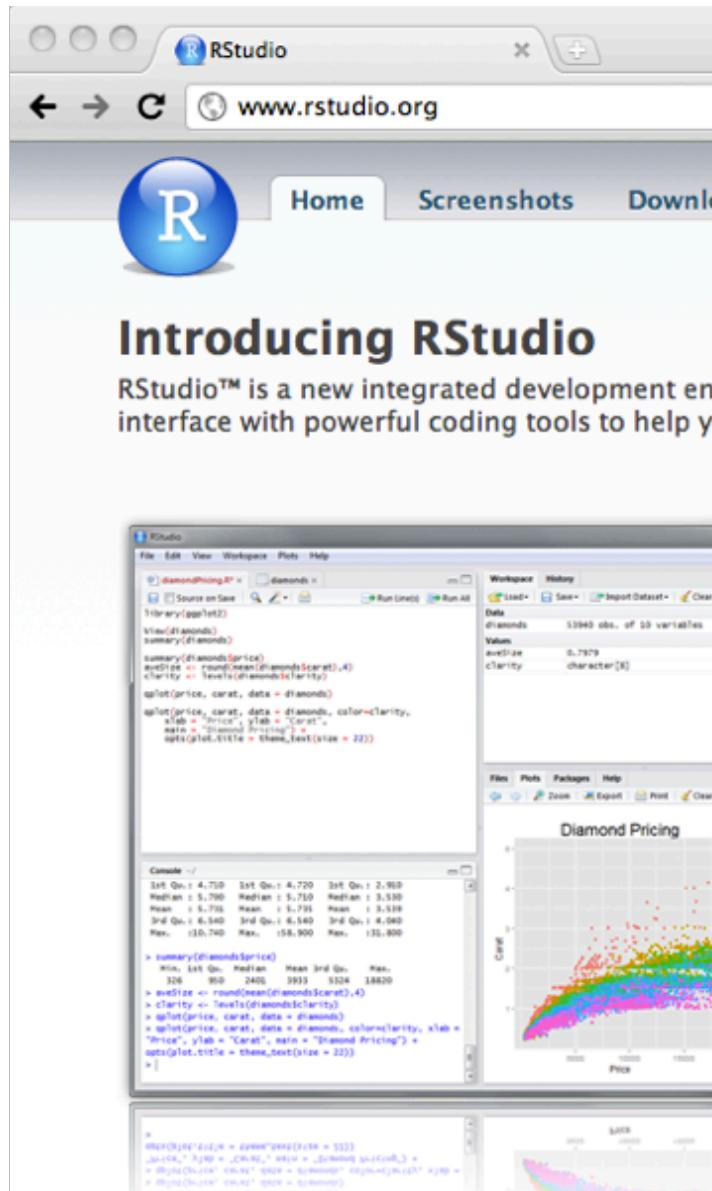
1. **Access:** It's free and runs on any operating system you are likely to have at home
 2. **Openness:** R is open source; there is a large community actively contributing to the codebase, and most practicing statisticians "share" their analyses through R programs
 3. **Expressiveness:** Far from a point-and-click interface, R is programmable and extendable; with it we will consider a range of complex data types



Computing: RStudio

This year we will also use a “front end” (formally an IDE or integrated development environment) for R called RStudio -- We do this because

1. **Access:** It's free and runs on any operating system you are likely to have at home
 2. **Uniformity:** As a program, R's native interfaces differ slightly from platform to platform -- RStudio gives us a uniform look so that you can move seamlessly from a Mac to a PC or even a Linux box
 3. **Ease of use:** Some of the non-statistical stumbling blocks for students in other terms involved knowing how to load data into R that was fetched from the web -- RStudio makes some of these routine operations easy



Computing

That said, any piece of digital technology develops and matures in a particular technological and social context; it's important to remember the following chain (which holds equally for MS Word, Google, Facebook or your favorite iPhone/Android App)

Software = Model of the world = An argument

This is especially true for a piece of “technical” software like R -- You will be learning how to **analyze a set of data**, with a mix of visual and numerical summaries, with a mix of simple mathematics, probability and simulation

Embedded in the commands you will type, in the small programs you will run, are a series of choices that the R designers have made -- It's sensible to ask why R looks and operates the way it does (are certain analyses easier to implement than others, for example)

About your assignments

This appeared in the New York Times a couple years ago; the main point of the article seems to be that R has a large user base (250K by one estimate) and that it's stealing significant market share from older programs like SAS

It reinforces why we chose R in the first place -- It has a lot of support from the statistics community and is **gaining traction among data analysts** outside this group

According to Daryl Pregibon (a statistician at Google and my old boss) "It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems."

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times **Business Computing**

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION

Search Technology Go

Inside Technology
Internet | Start-Ups | Business Computing | Companies

Data Analysts Captivated by R's Power



Stuart Issett for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE
Published: January 6, 2009

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as [Google](#), [Pfizer](#), [Merck](#), [Bank of America](#), the InterContinental Hotels Group and Shell use it.

E-MAIL | PRINT | SINGLE PAGE | REPRINTS | SAVE | SHARE

ARTICLE TOOLS SPONSORED BY
THE WRESTLER
3 GOLDEN GLOBE NOMS

Our approach

Your text takes a slightly tedious formula-laden approach to the field, one that I find a little disheartening -- We will use the book as a reasonable reference, but push the concepts much much farther, bringing this class of ye olde statistics

We will also emphasize re-randomization (permutation) and resampling as basic tools for statistical inference rather than classical tests and intervals -- We will show how these newer developments agree with the analytical approach taken by your text

Next, your assignments will be much less about formula plug-and-go, but instead emphasize reasoning and analysis -- You will not not “scratch anything down” on paper in this class

You will, however, be asked to compute a fair bit -- So whether it's with R or some other computing platform, you need to be prepared to work with data (computation, visualization, modeling)

Finally, whenever possible we will resort to a geometric view of statistics rather than something analytical (this will make more sense as we go) -- The bottom line is that hopefully it will be better tuned to the needs of an engineer



08:50:42



09: 14:45

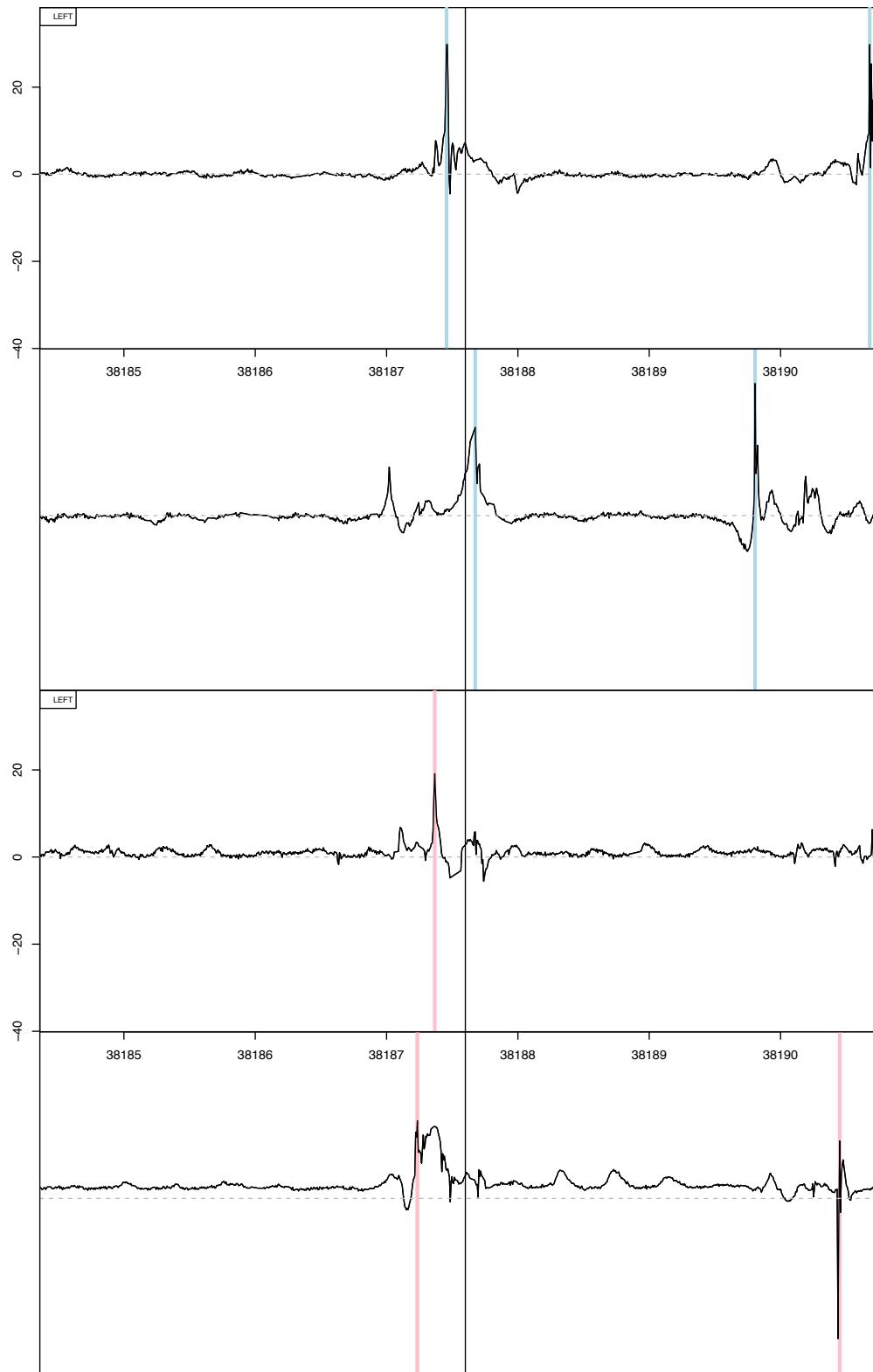




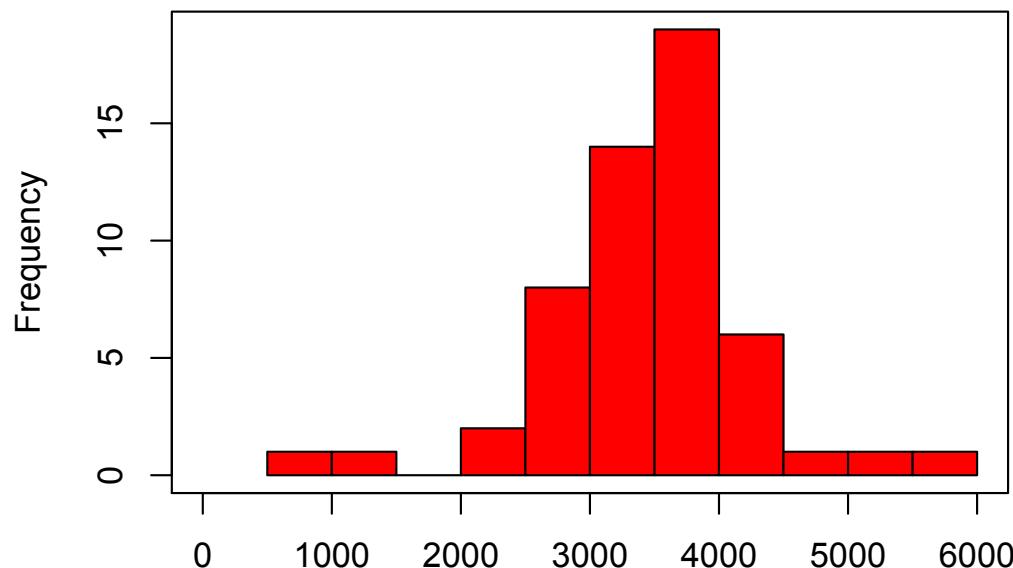
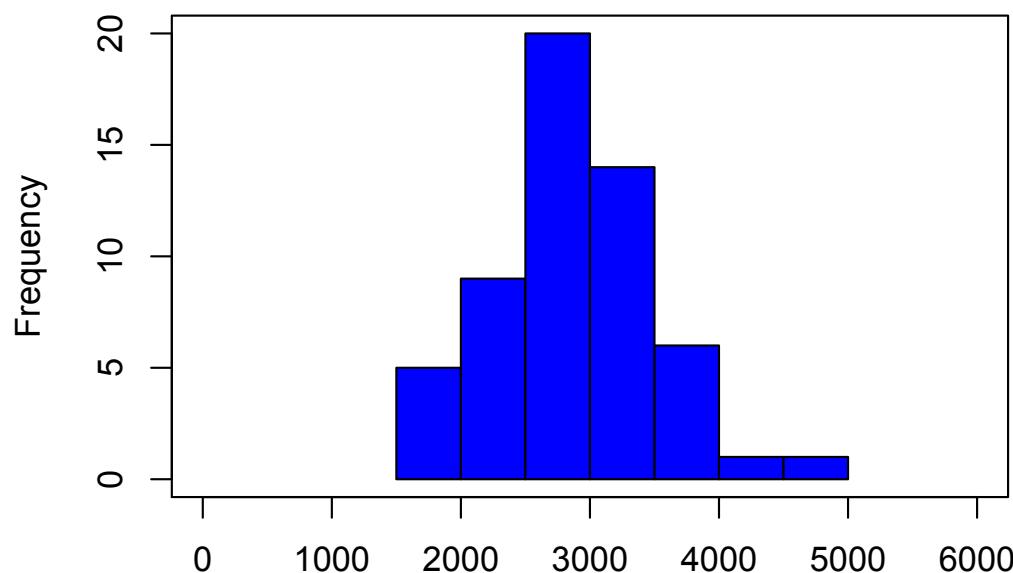




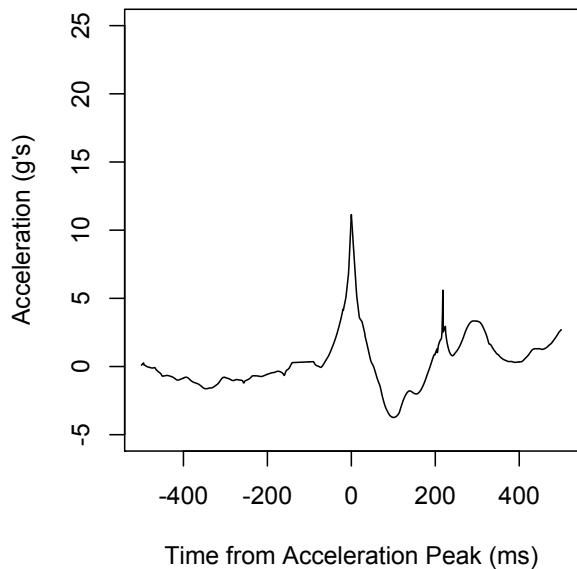
TCR. 07:36. 27:14



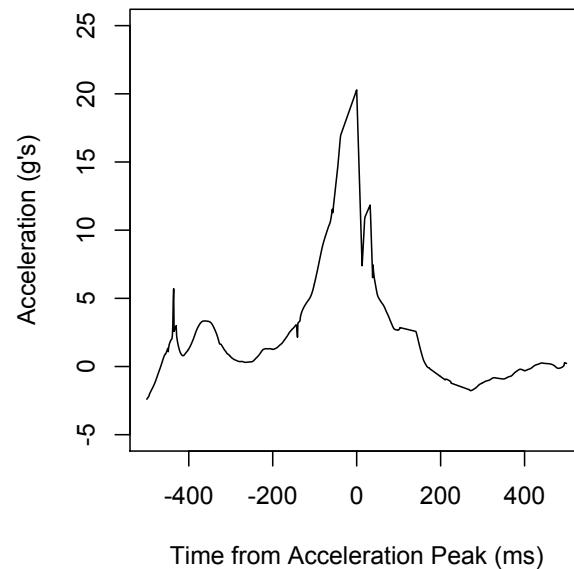
Force (N) Comparison



Garcia Non-Punch



Garcia Punch

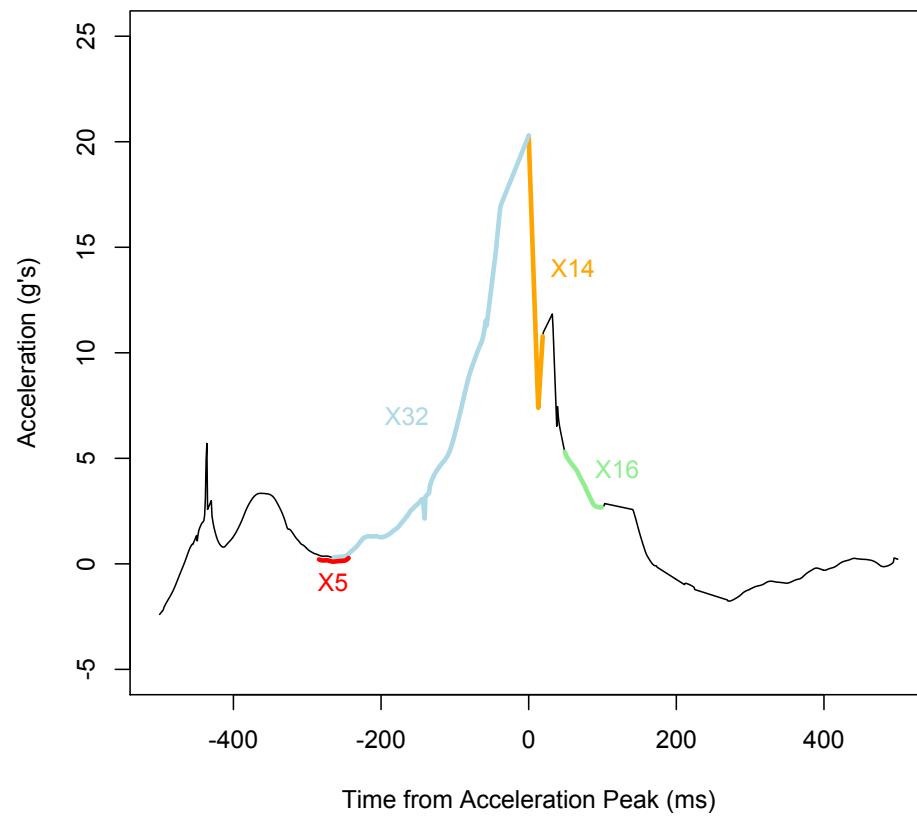


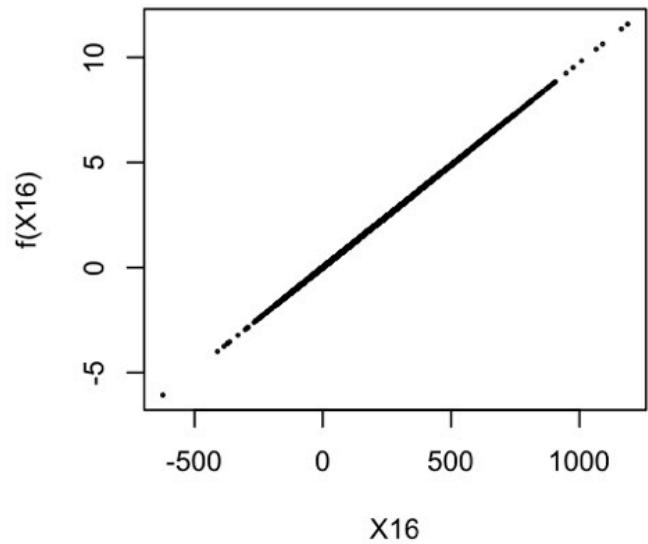
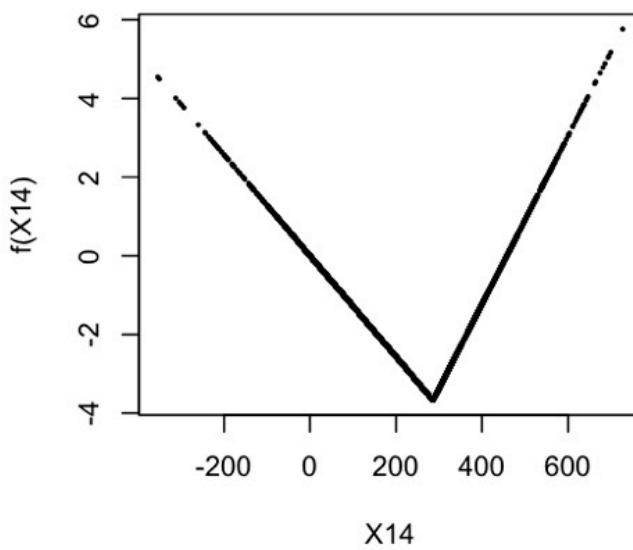
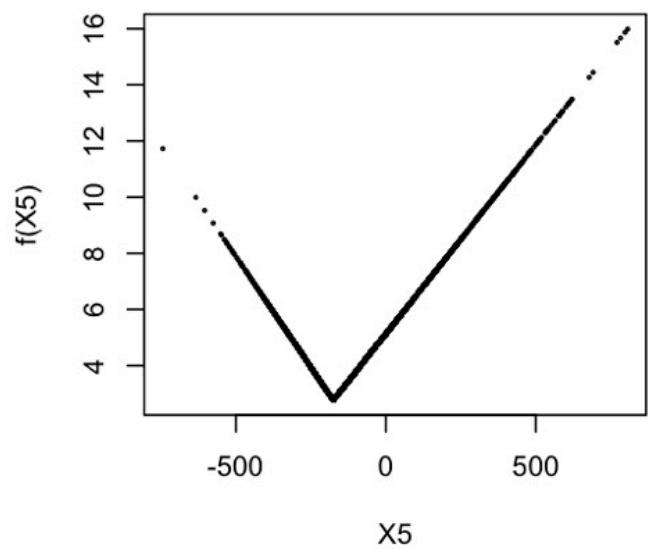






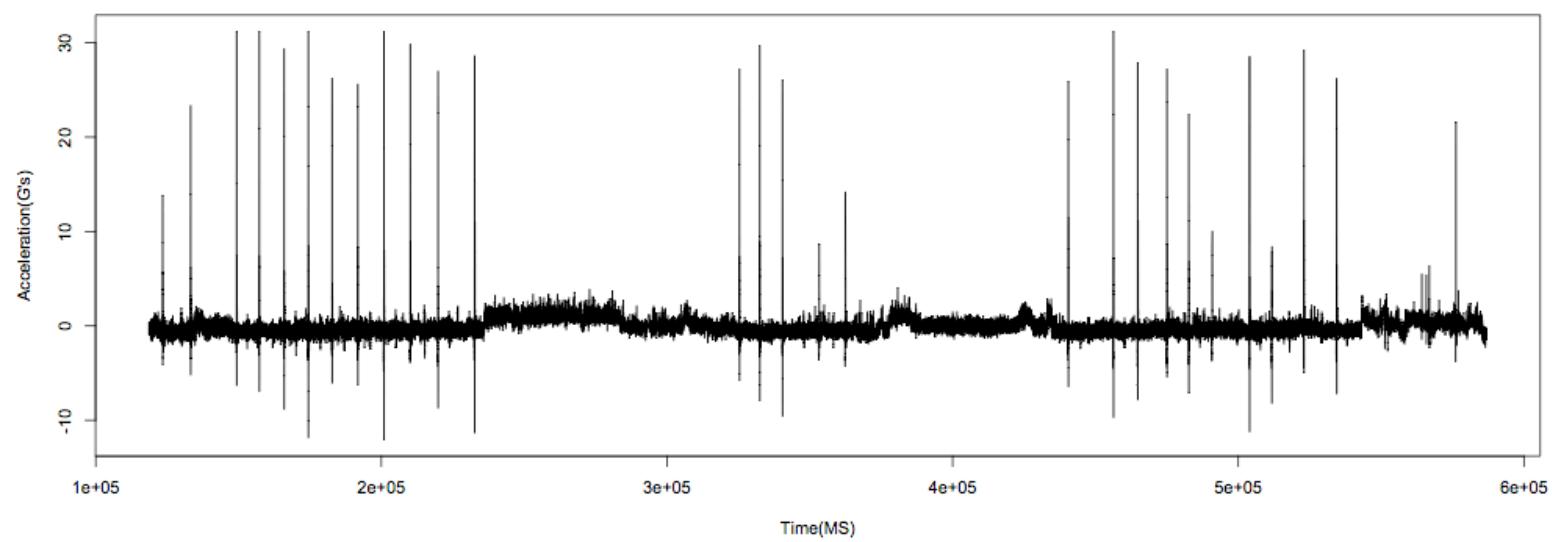
Garcia Punch

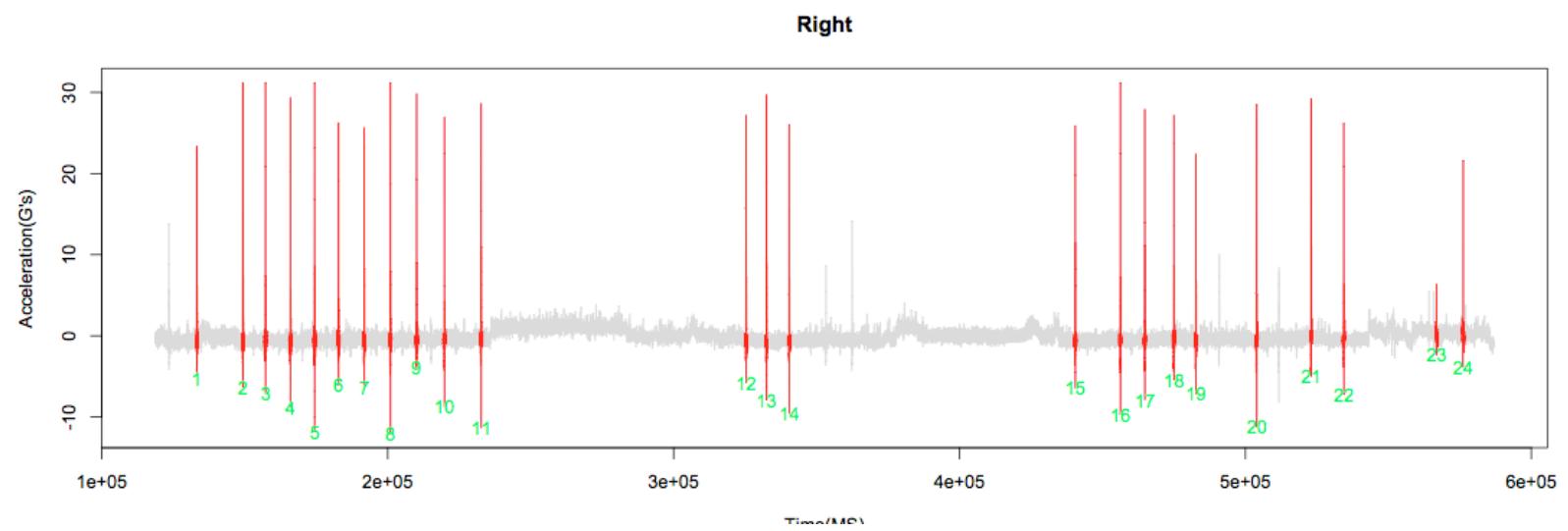


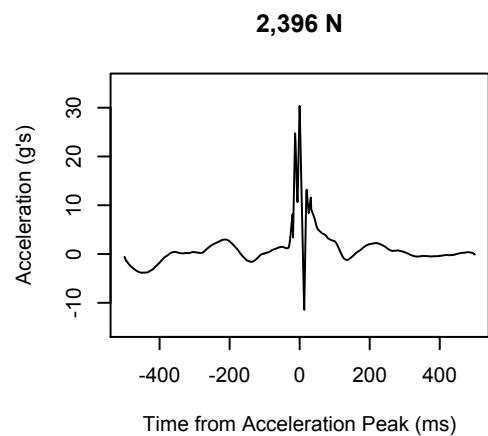
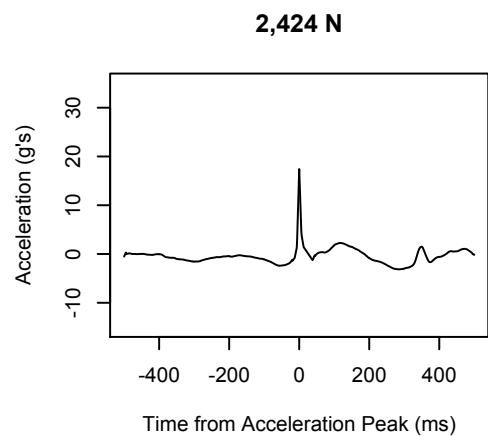
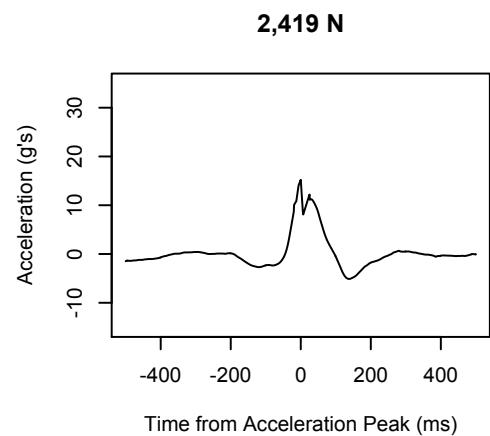
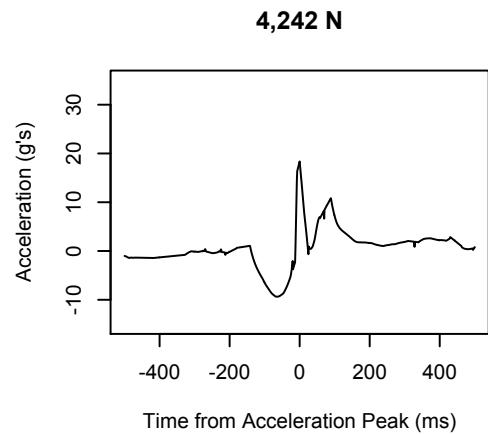
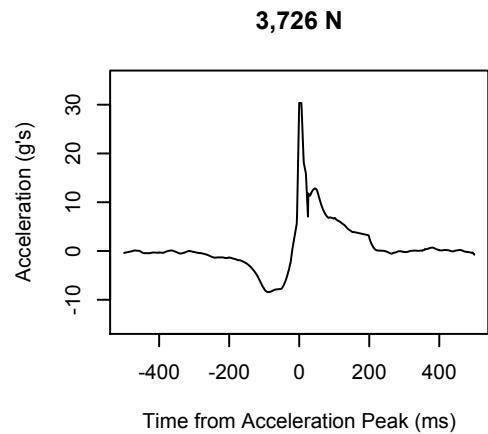
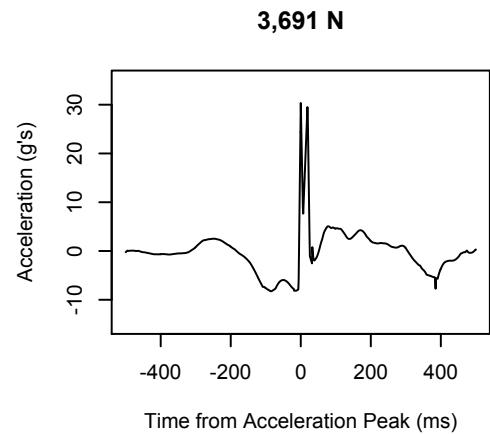


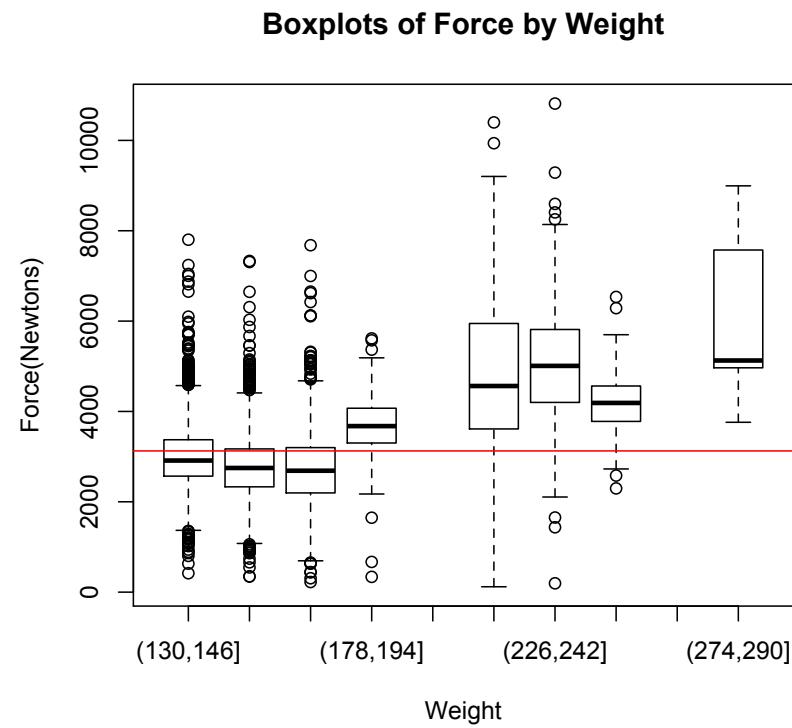
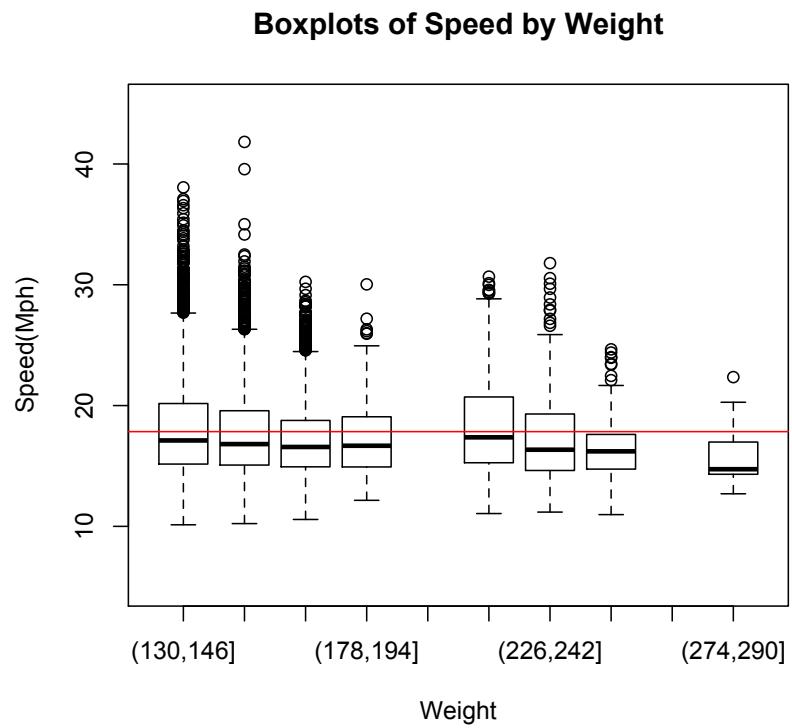


Right



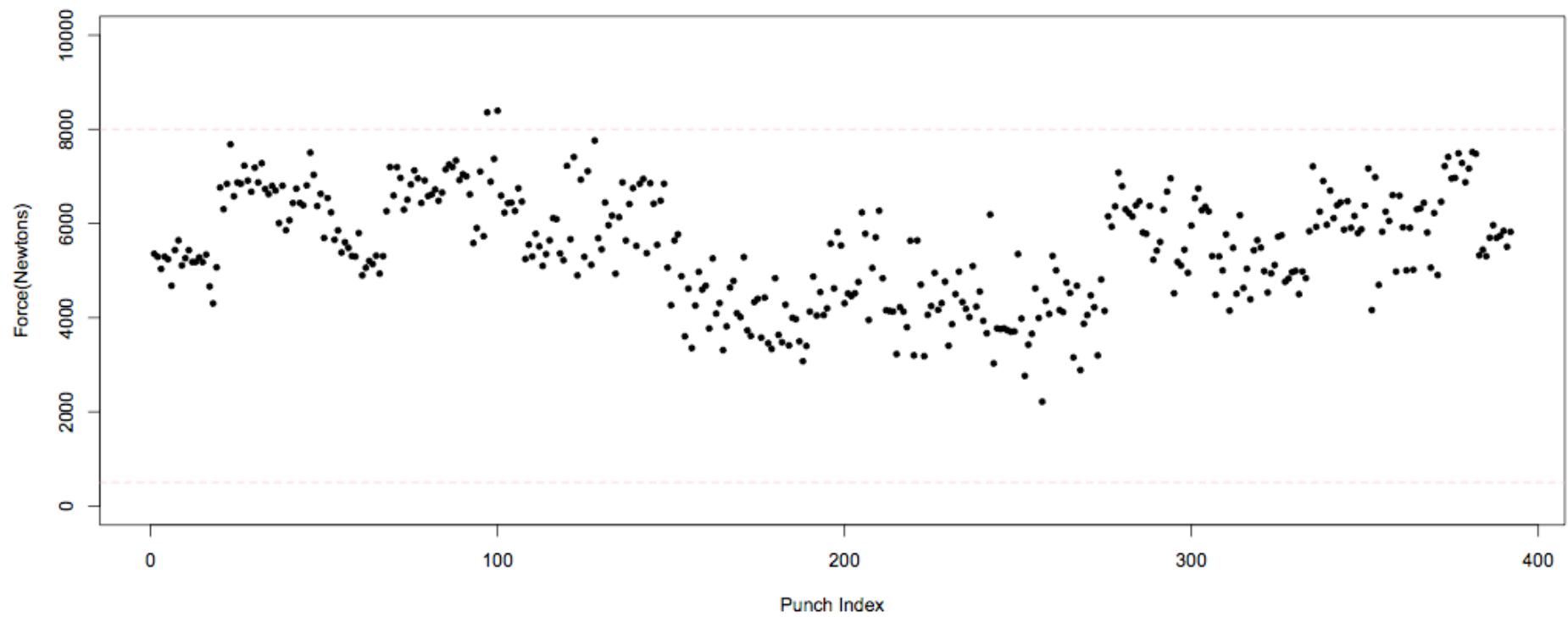


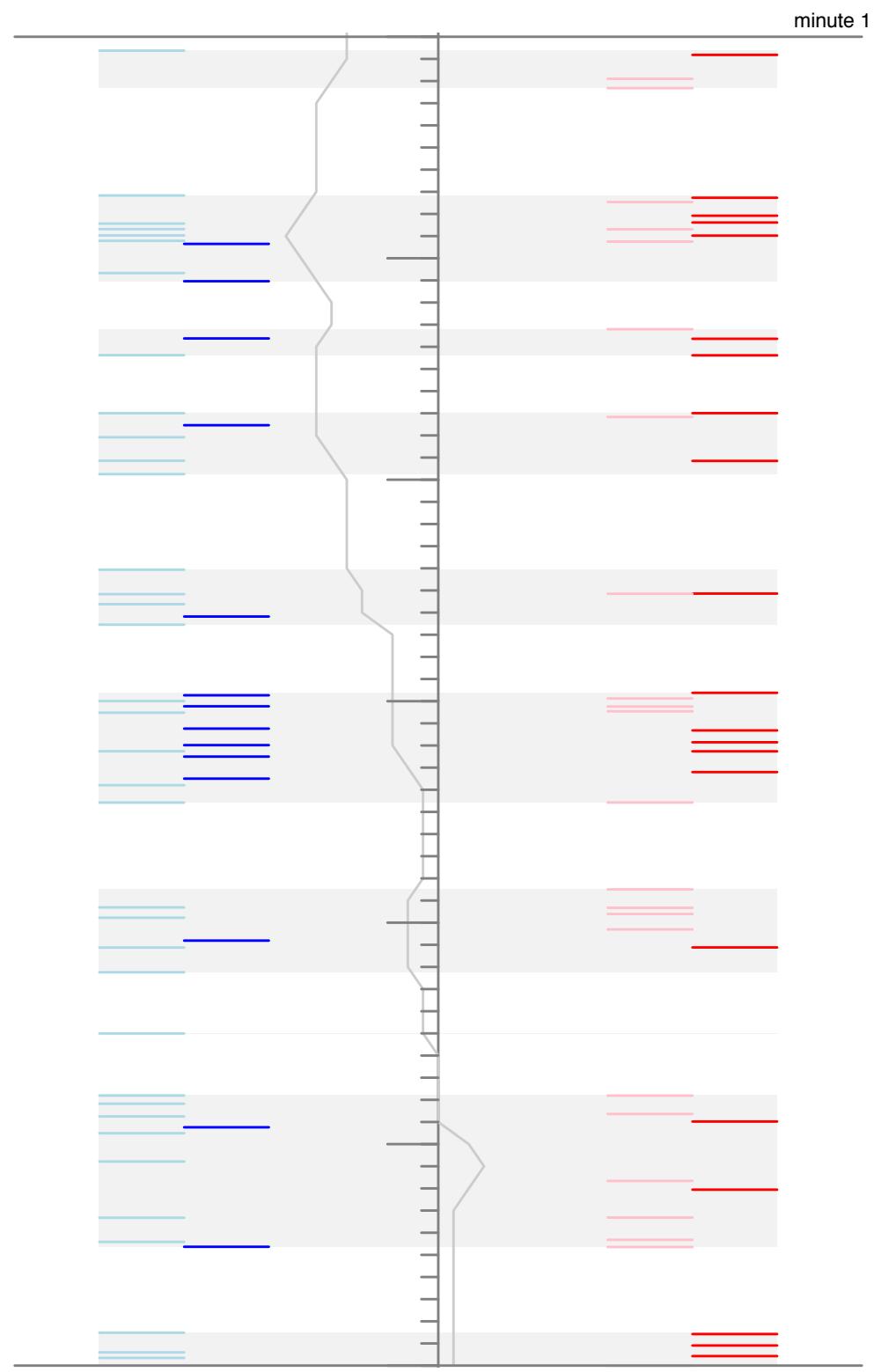


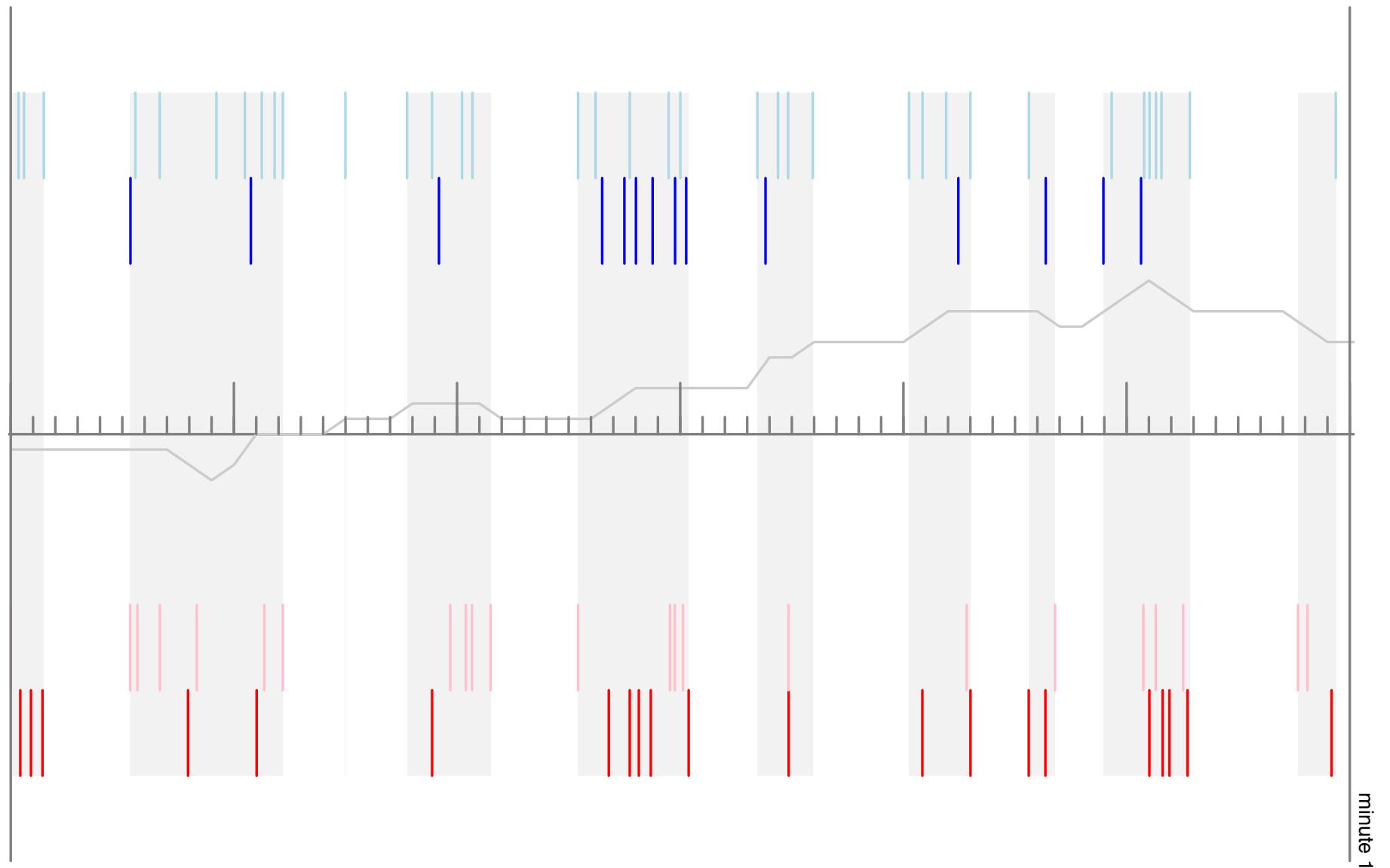


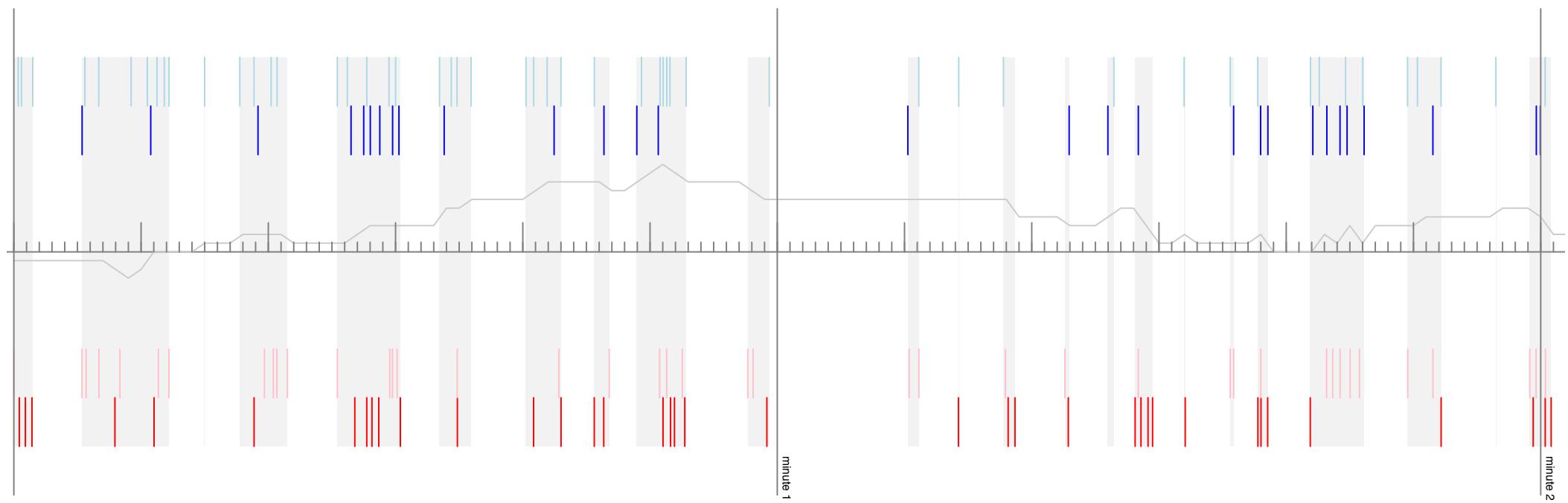
```
<?xml version="1.0" encoding="UTF-8"?>
<splineModel>
  <term>
    <coefficient>5817.4512456</coefficient>
    <se>2734.527438</se>
    <expression>5817.4512456</expression>
  </term>
  <term>
    <coefficient>365.2001095</coefficient>
    <se>32.998199</se>
    <product>
      <variable>
        <name>forearm.hit</name>
      </variable>
    </product>
    <expression>365.2001095*x[0]</expression>
  </term>
  <term>
    <coefficient>-423.9208580</coefficient>
    <se>77.889330</se>
    <product>
      <variable>
        <name>commitment</name>
      </variable>
    </product>
    <expression>-423.9208580*x[1]</expression>
  </term>
  <term>
    <coefficient>-682.2593177</coefficient>
    <se>58.719188</se>
    <product>
      <variable>
        <name>armweight.hit</name>
      </variable>
    </product>
```

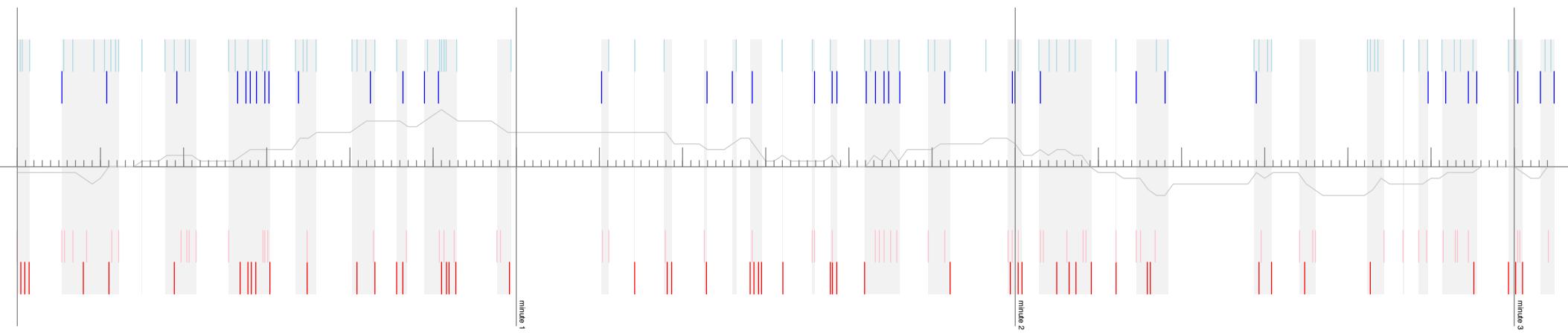
PunchForce Test 01/05/2011 - Heavyweight Boxer Predicted Punch Force















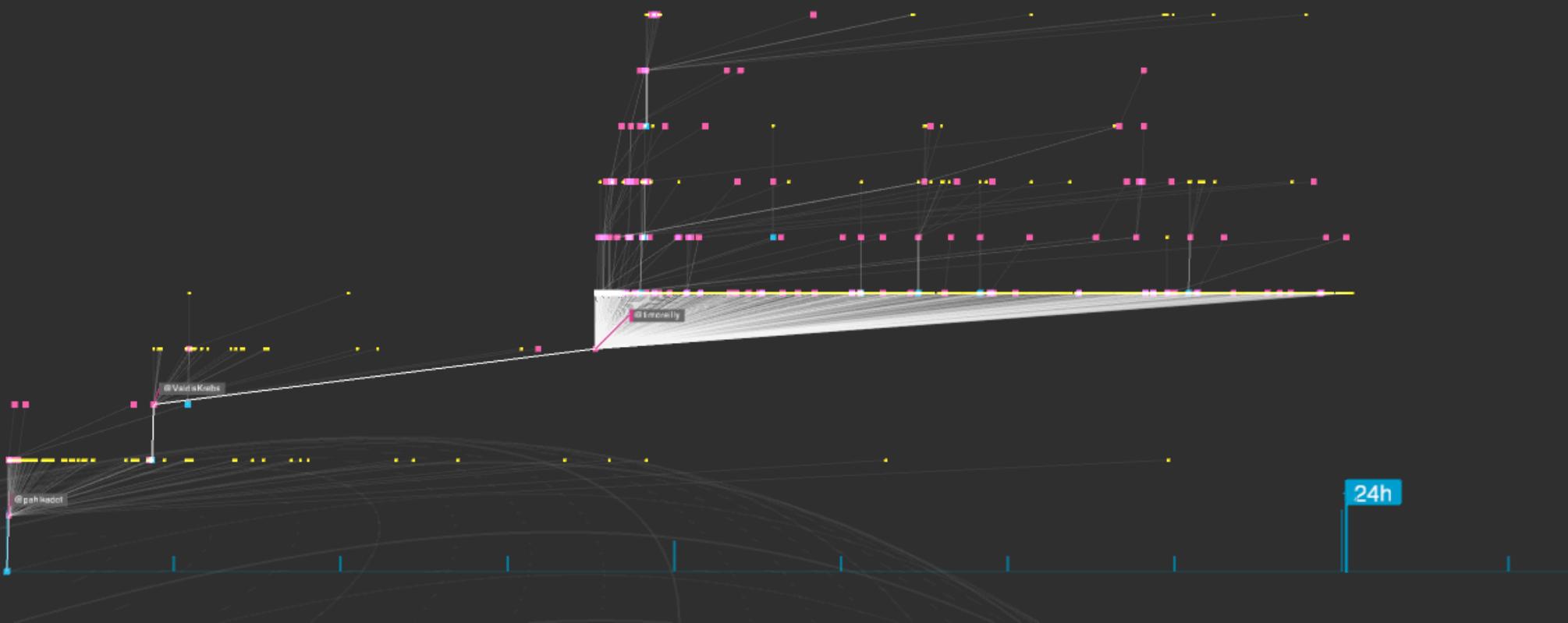
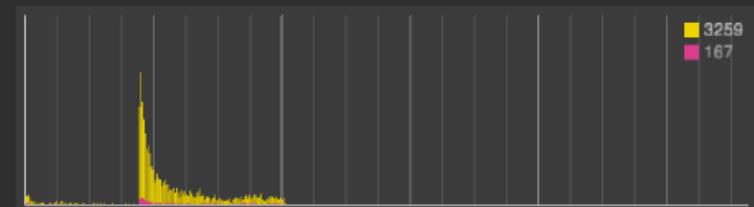




OP-ED COLUMNIST; America Goes Dark

By PAUL KRUGMAN

Tue Aug 10 11:35:43 EDT 2010

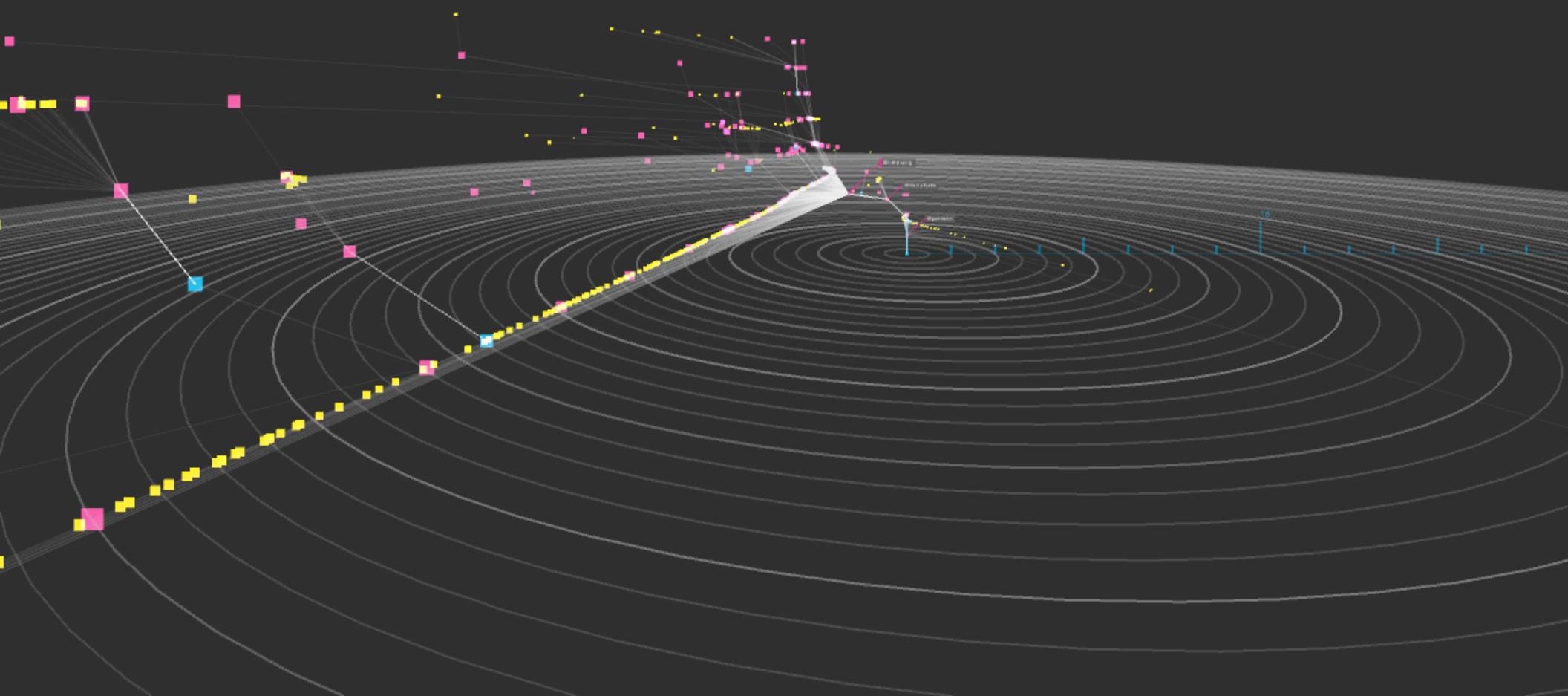




OP-ED COLUMNIST; America Goes Dark

By PAUL KRUGMAN

Wed Aug 11 16:21:15 EDT 2010

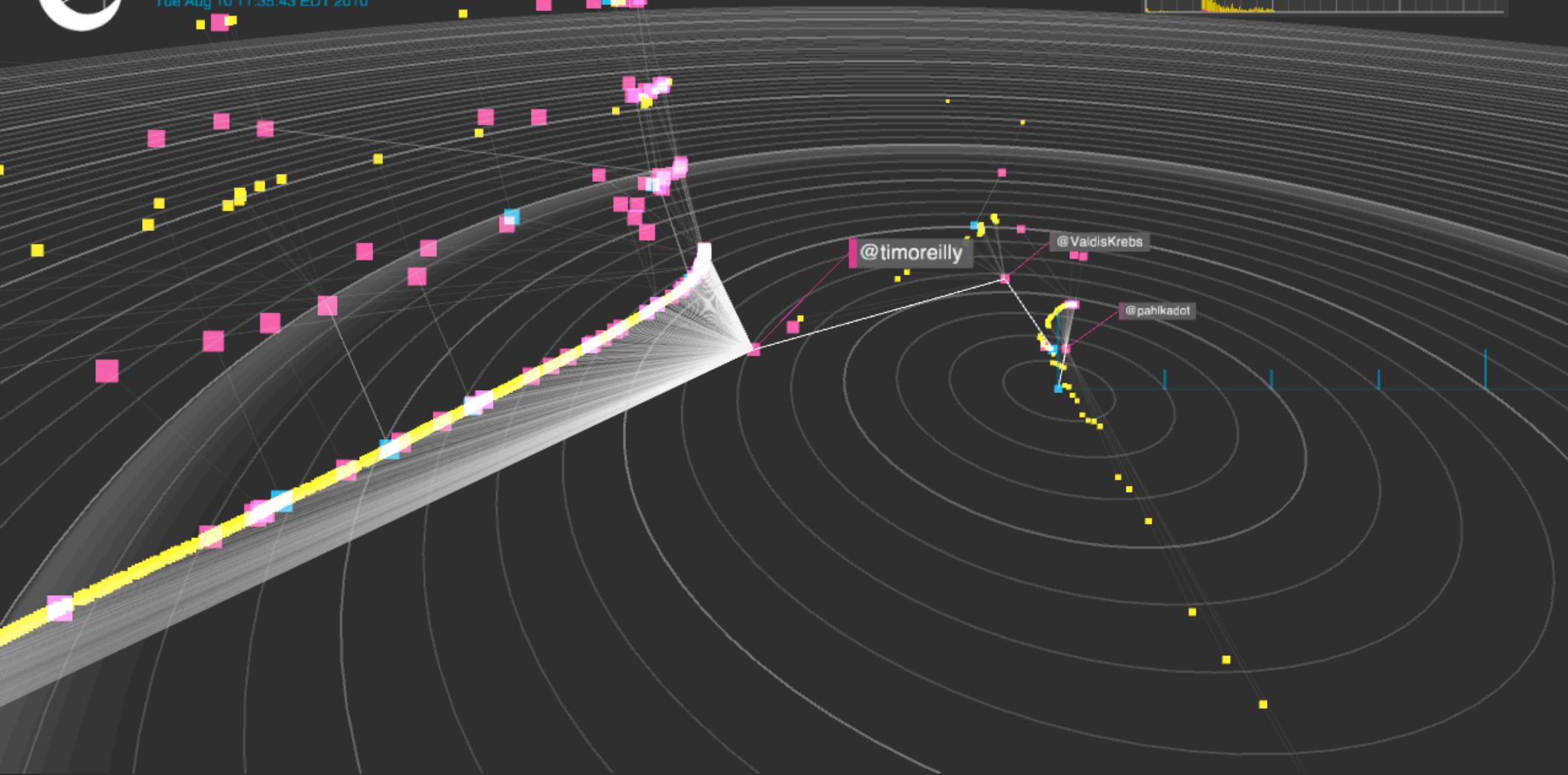




OP-ED COLUMNIST; America Goes Dark

By PAUL KRUGMAN

Tue Aug 10 11:35:43 EDT 2010





OP-ED COLUMNIST; America Goes Dark

By PAUL KRUGMAN

Tue Aug 10 11:35:43 EDT 2010



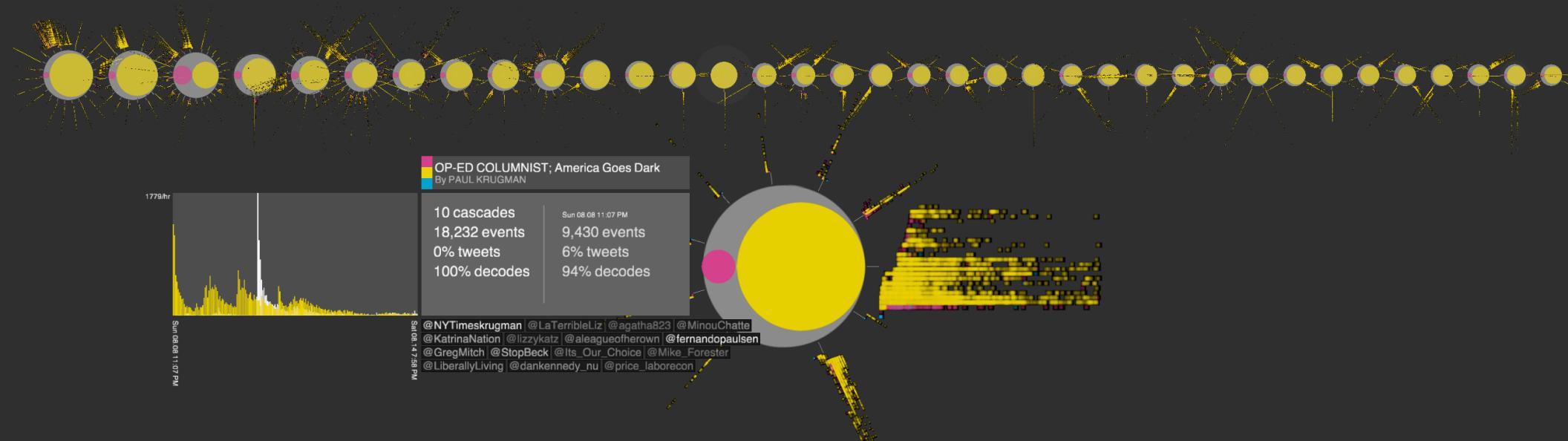
America Goes Dark: a frightening and all-too-possible wrong turn.
<http://nyti.ms/9qsmMW>

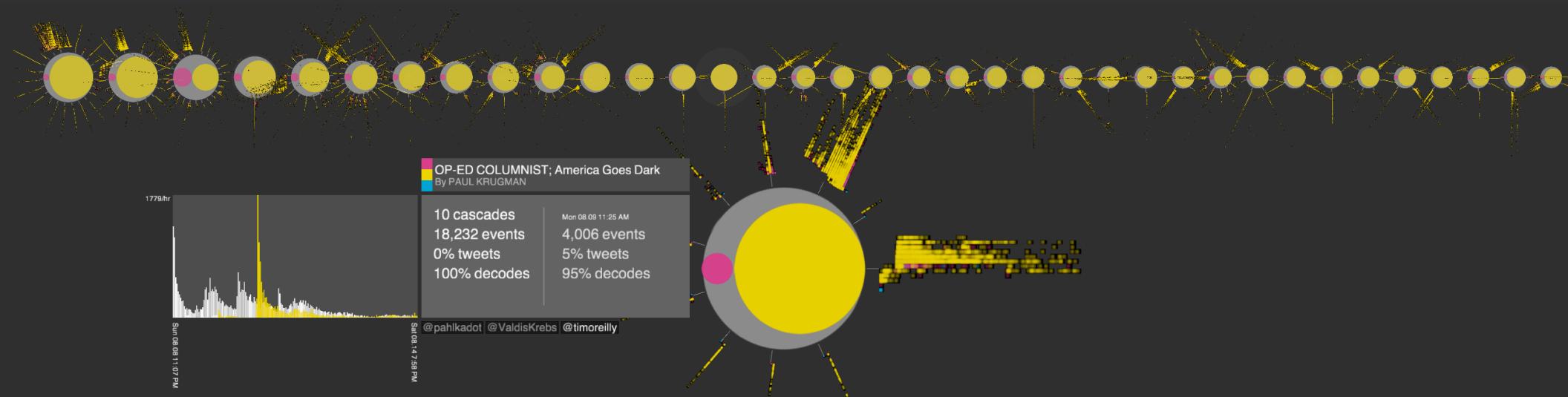
@timoreilly

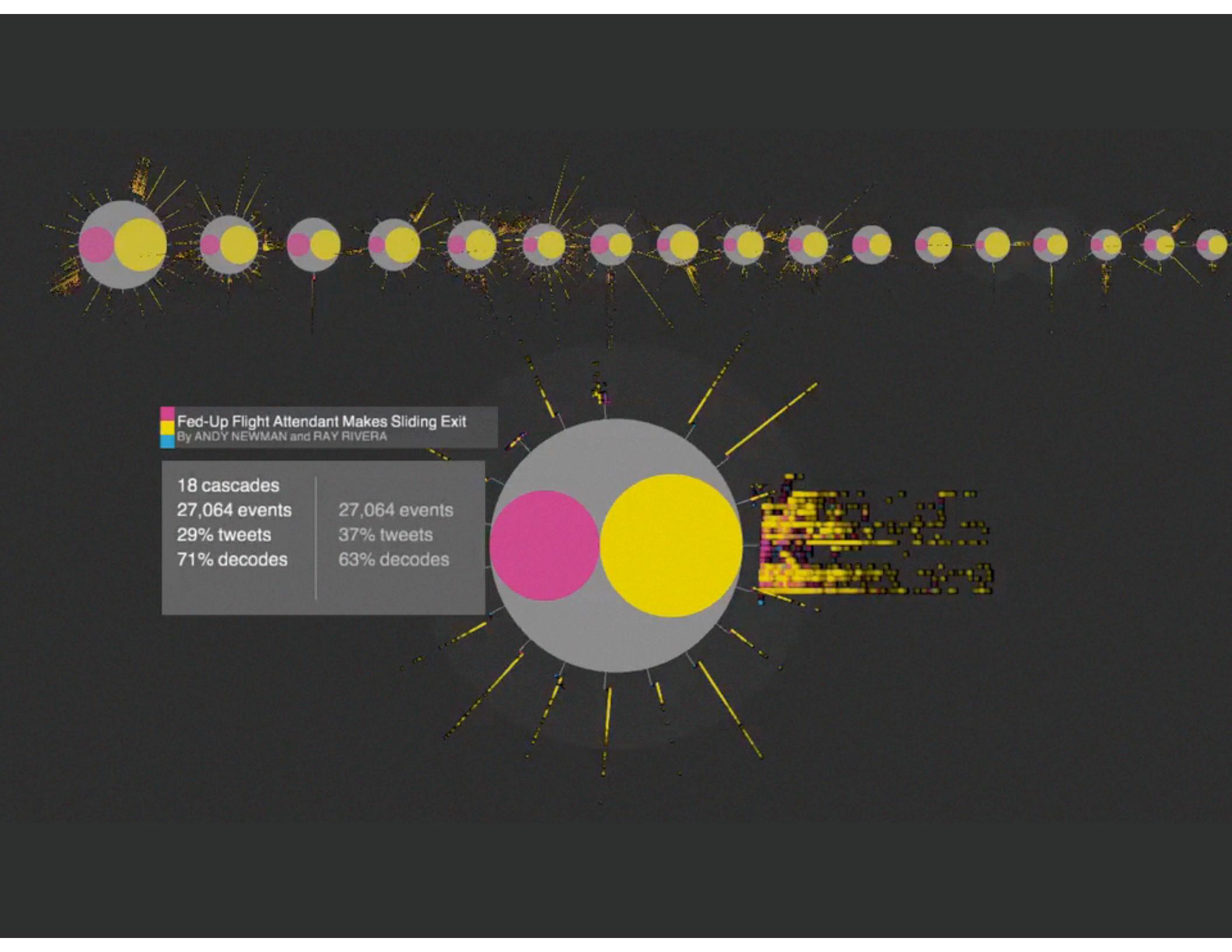
1,426,193 followers

Sebastopol, CA

235.0







Fed-Up Flight Attendant Makes Sliding Exit
By ANDY NEWMAN and RAY RIVERA

18 cascades
27,064 events
29% tweets
71% decodes

27,064 events
37% tweets
63% decodes

