Data Athletics

Analyzing triathlete's performance - Abhi

Agenda

- Project background
- Hypothesis
- Data analysis and Visualization
- Code
- Future
- Questions

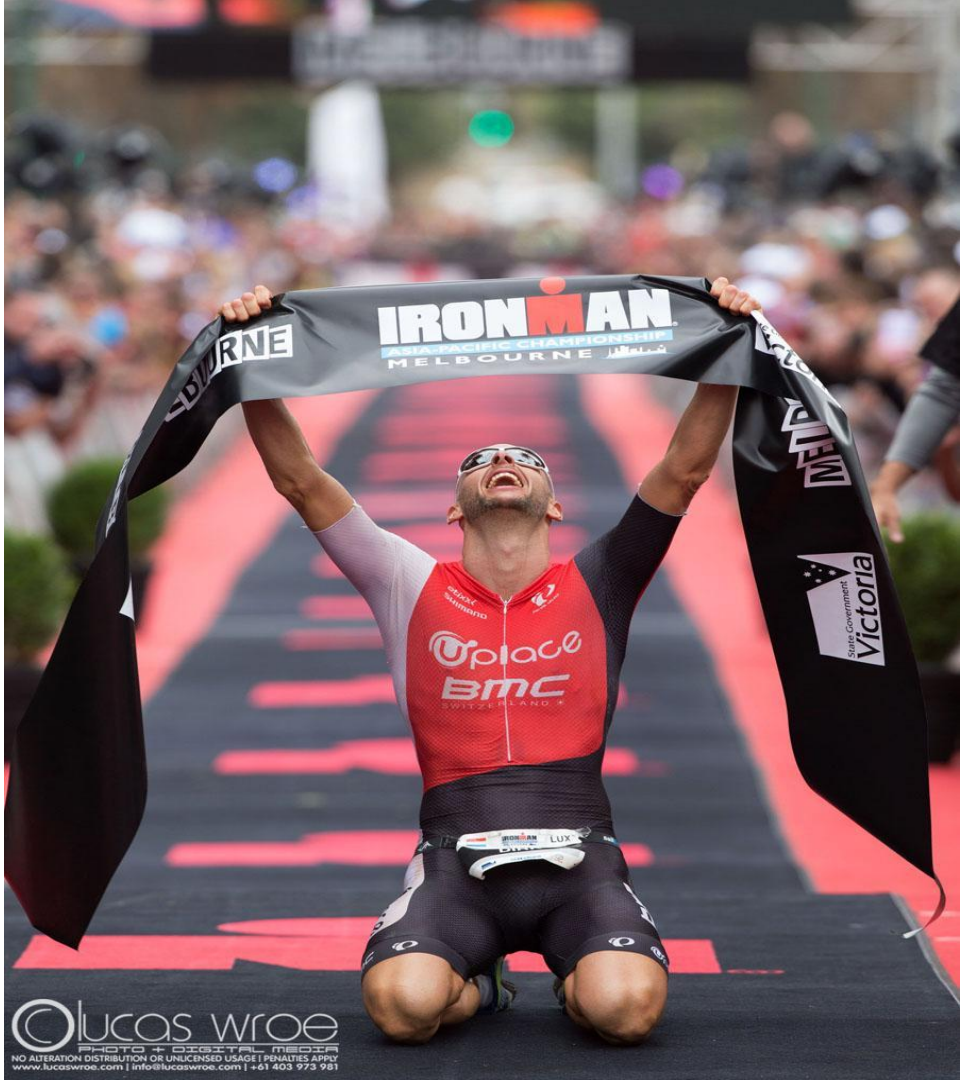
Hypothesis

- Can we predict athlete’s future performances based on race history and training data ?
- Looking back can we quantify what went well and what did not during a race ?
- All (most) serious athletes are data geeks and wealthy

(Half)IronMan Triathlon

- Born on the Hawaiian island of Oahu in 1978
- Ironman: 2.4-mile swim, 112-mile bike and 26.2-mile
  Half Ironman: 1.2 mile swim, 56 mile bike, 13.1 mile run
- Over the years, 88 different countries have been represented at the championship
- More than 60,000 triathletes try each year, but only about 2000+ win coveted Kona Slots (1700+ qualifiers, plus lottery winners, etc.)
- More than +$150 million in revenue


603
339


IRONMAN
ASIA-PACIFIC CHAMPIONSHIP
MELBOURNE
Victoria
Uplace
BMC
lucas wroe
PHOTO + DIGITAL MEDIA
NO ALTERATION DISTRIBUTION OR UNLICENSED USAGE | PENALTIES APPLY
www.lucaswroe.com | info@lucaswroe.com | +61 403 973 981

Data Gathering

- Nodejs + Cheerio for scraping
- Weather data
- Amazon RDS
- ~400K rows for Half Ironman
- ~450K rows for Ironman
- ~12 Race factors (Attributes)

Challenges:

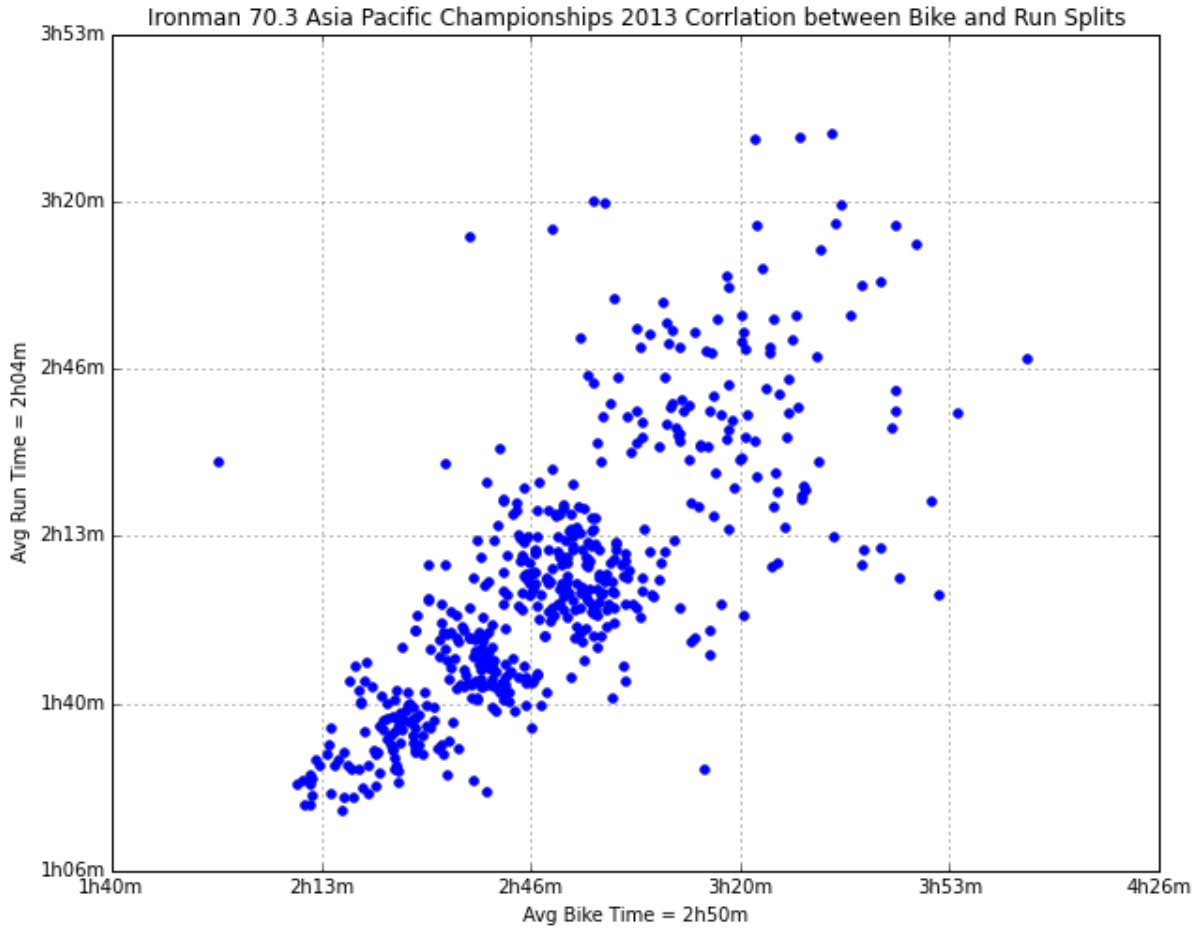
- No clean API
- Unreliable website (Athlinks)
- Missing data

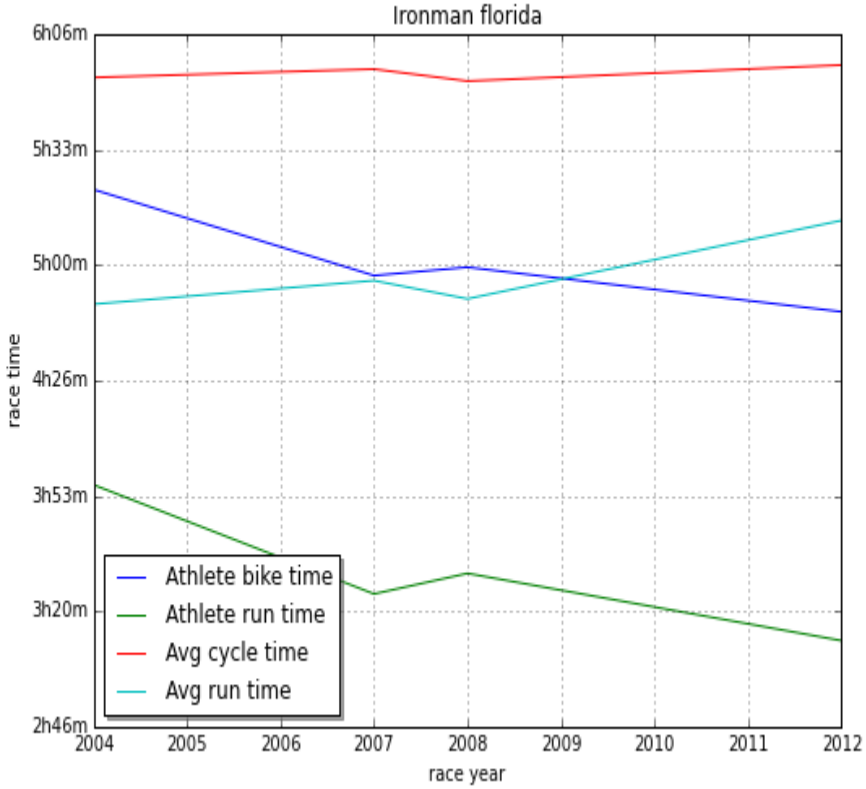
Dataset

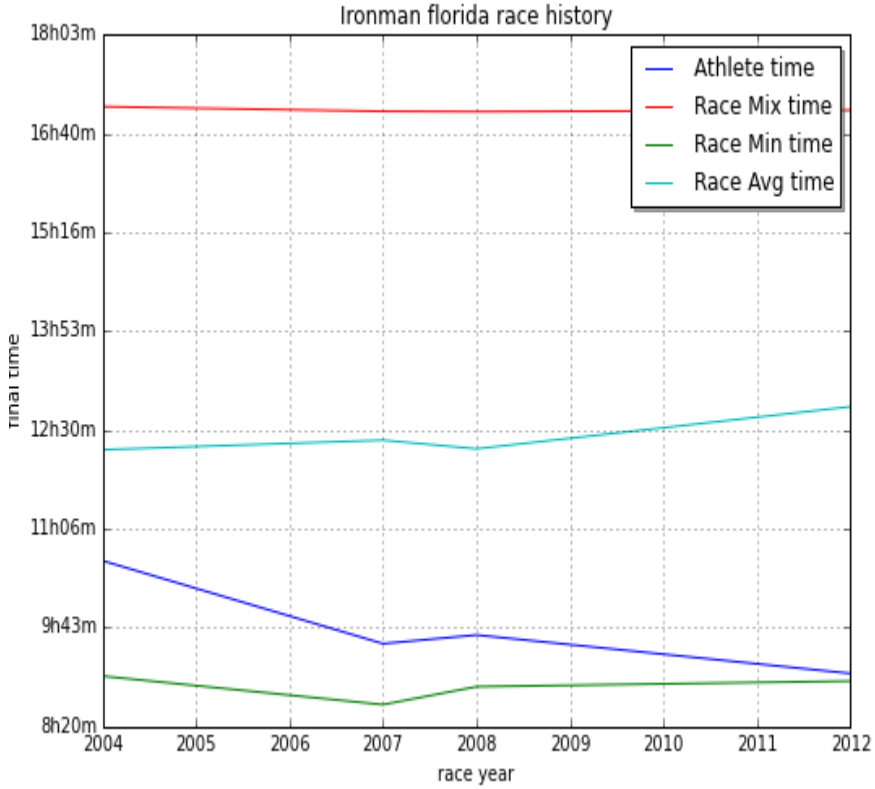
- Race result data comprised of Bike, Run, Swim and Final time

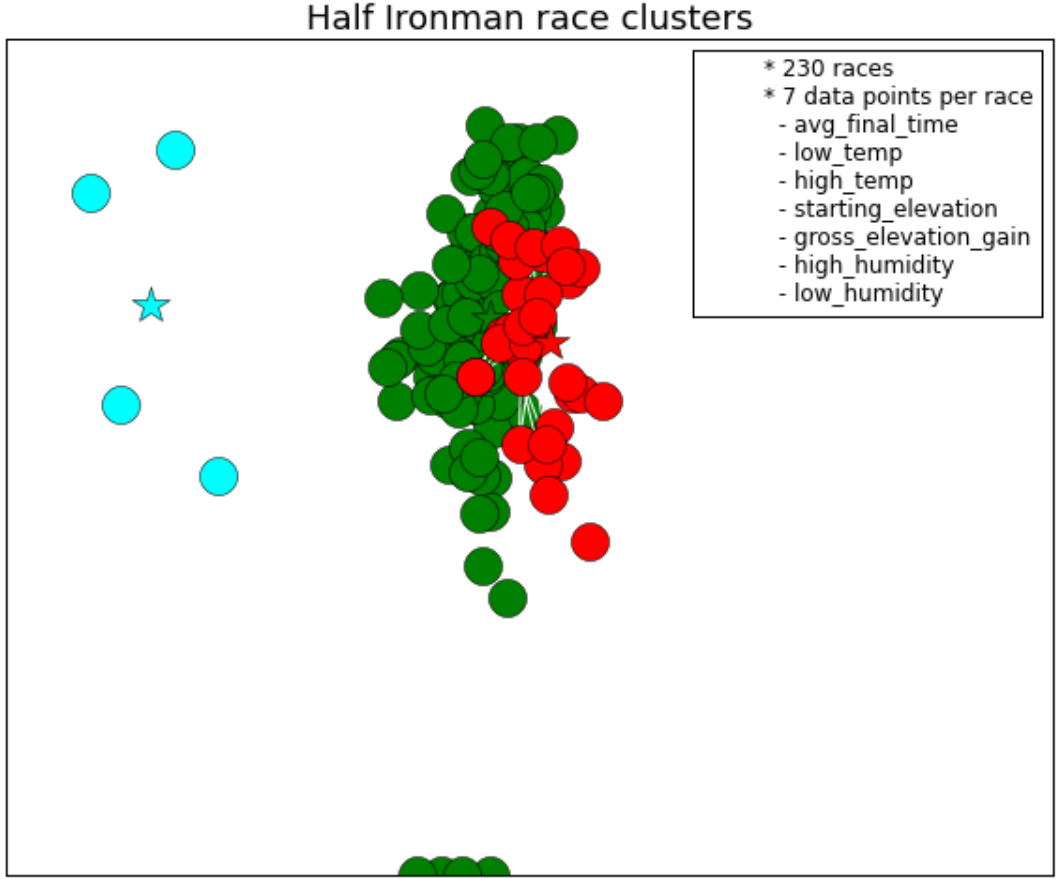
- Race factor data comprised of Starting/Max and Gross elevation, Temperature, Humidity, Peak wind speed, Sunrise and Sunset etc

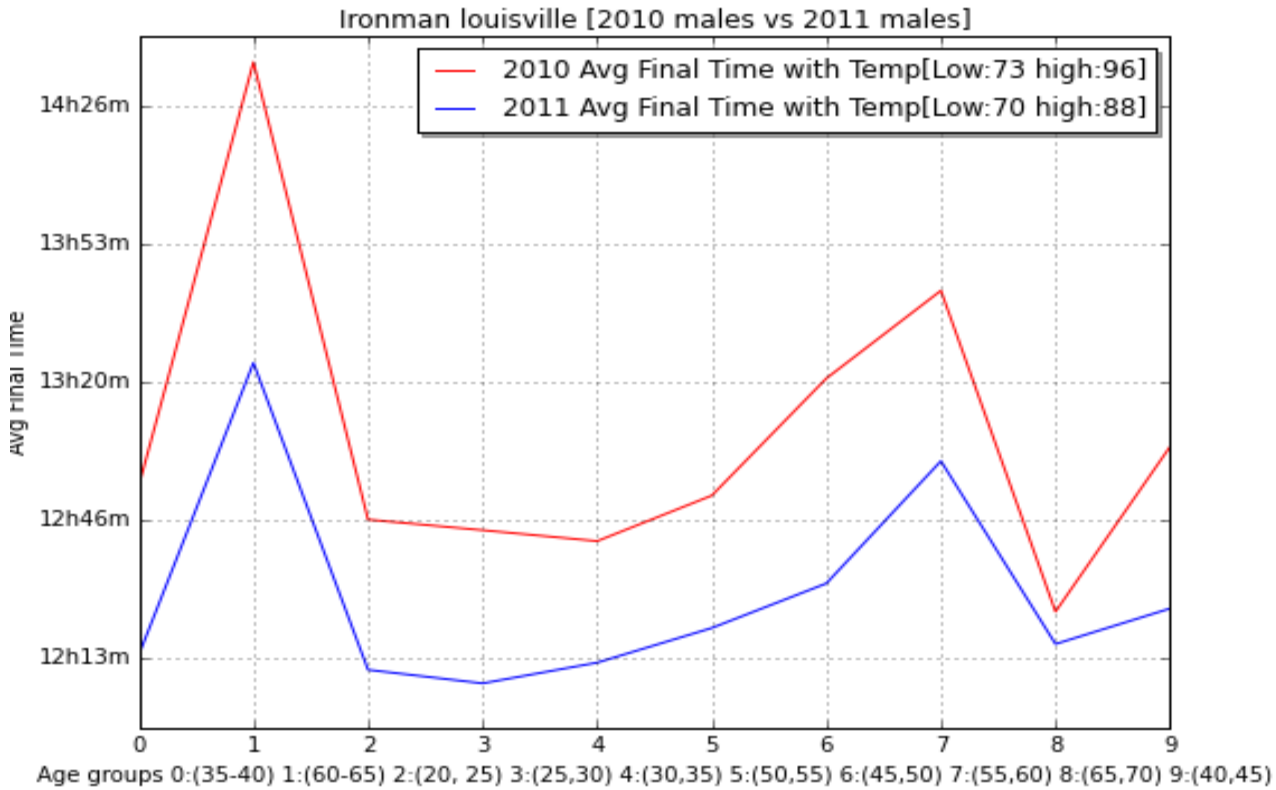
Data Explo

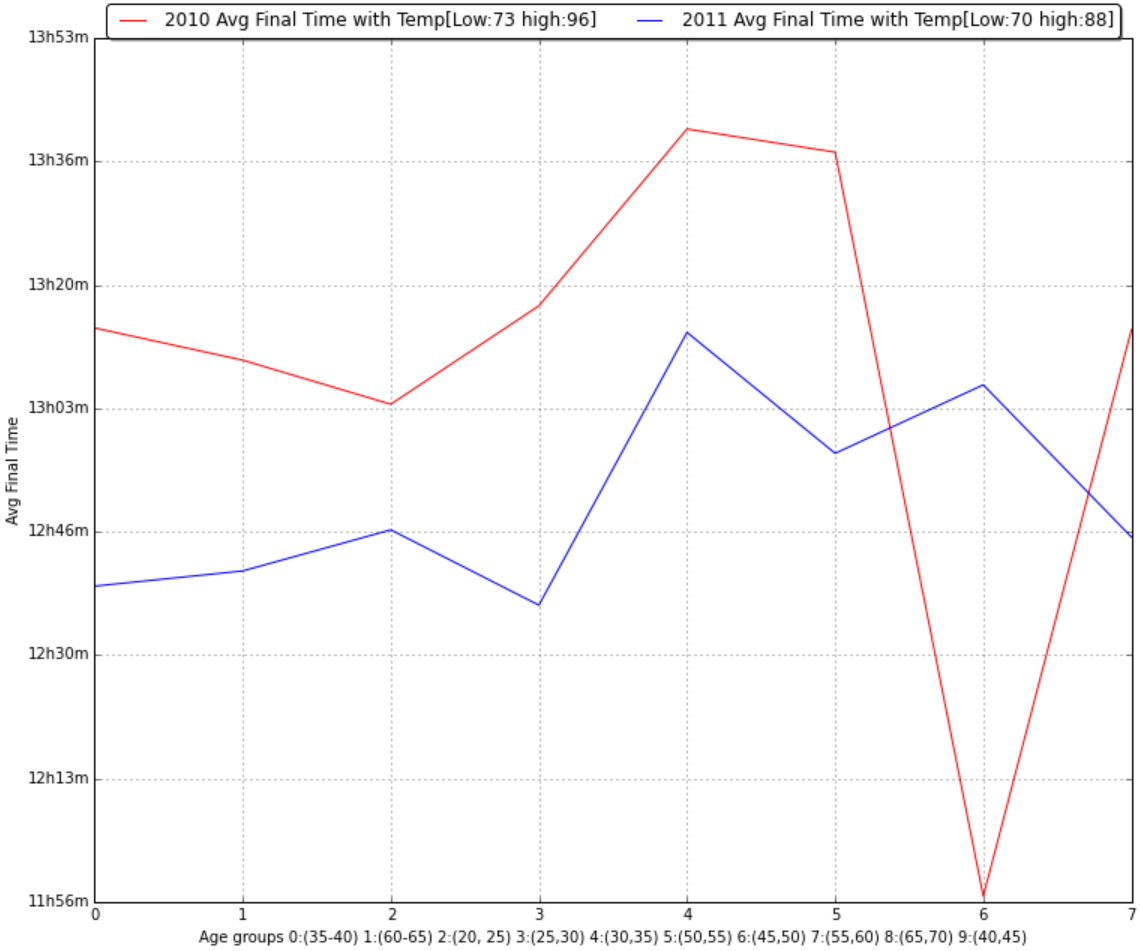

Ironman 70.3 Asia Pacific Championships 2013 Corrlation between Bike and Run Splits
3h53m
3h20m
2h46m
2h13m
1h40m
1h06m
Avg Run Time = 2h04m
1h40m
2h13m
2h46m
3h20m
3h53m
4h26m
Avg Bike Time = 2h50m


Ironman florida
6h06m
5h33m
5h00m
4h26m
3h53m
3h20m
2h46m
race time
2004
2005
2006
2007
2008
2009
2010
2011
2012
race year
Athlete bike time
Athlete run time
Avg cycle time
Avg run time


Ironman florida race history
18h03m
16h40m
15h16m
13h53m
12h30m
11h06m
9h43m
8h20m
final time
2004
2005
2006
2007
2008
2009
2010
2011
2012
race year
Athlete time
Race Mix time
Race Min time
Race Avg time


Half Ironman race clusters
* 230 races
* 7 data points per race
- avg_final_time
- low_temp
- high_temp
- starting_elevation
- gross_elevation_gain
- high_humidity
- low_humidity


Ironman louisville [2010 males vs 2011 males]
2010 Avg Final Time with Temp[Low:73 high:96]
2011 Avg Final Time with Temp[Low:70 high:88]
14h26m
13h53m
13h20m
12h46m
12h13m
Avg final time
0
1
2
3
4
5
6
7
8
9
Age groups 0:(35-40) 1:(60-65) 2:(20, 25) 3:(25,30) 4:(30,35) 5:(50,55) 6:(45,50) 7:(55,60) 8:(65,70) 9:(40,45)


2010 Avg Final Time with Temp[Low:73 high:96]
2011 Avg Final Time with Temp[Low:70 high:88]
13h53m
13h36m
13h20m
13h03m
12h46m
12h30m
12h13m
11h56m
Avg Final Time
0
1
2
3
4
5
6
7
Age groups 0:(35-40) 1:(60-65) 2:(20, 25) 3:(25,30) 4:(30,35) 5:(50,55) 6:(45,50) 7:(55,60) 8:(65,70) 9:(40,45)

Neural Network

Feed Forward Network
Pros:

- Predict race distribution for a given set of race factors

- Confirmed hypotheses about race factors affecting population in a consistent way

Cons:

- Wasn’t useful for personalized results

Code setup

- 6 Input (Race factors) and 15 output (race bins with population percent) layers with 5 hidden layers (with Sigmoid activation function)
- 90/10 split into Train/Test dataset
- 1000 epochs took ~45mins
- Standard sum-of-squares error
- Used Pybrain python library
- Code on Github => Feed forward network

Pivot

Power Analysis

- Cyclists: The power a rider would need to produce to achieve a certain speed or event time or conversely the speed or time that could be expected for a given power output.

Can we do similar for running ?

Physiological Models

Modeling the energetics of 100-m running by using speed curves of world champions - Laurent M Arsac and Elio Locatelli

Mathematical analysis of running performance and world running records

- Francois Pfironnet and Guy Thibault

The energetics of endurance running

- di Prampero PE, Atchou G, Brückner JC, Moia C.

Power analysis

Power (W/kg)=(C * n*V - (C * n*V * (.5 * (V * 8.33-1)))) + C_aero + C_kin

Datathletics Power Calculator - ***Power*** *is the rate of doing work*

Careo = Energy required to overcome air resistance

where:

C_aero = k * d^3 * n ^-1 * t^-3)

k = 0.5*p*Af*Cd
p = sea level air density
Af = frontal_area
Cd = drag coefficient = 0.9
n = 0.5

V = avg running velocity = distance/time
Ckin = Energy required to change kinetic energy

C_kin = 0.5 * n^-1 * d^2 * t^-3

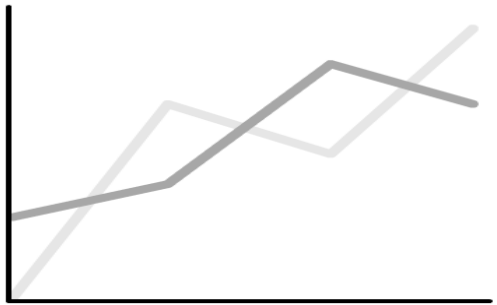
Next Steps ?

- Power curve
- Test with athletes
- Get Feedback
- Continue data exploration

Datathletics

Runs | Power Curve | Profile | Blog | Log Out

Alex's Power Curve



Alex Pape
Alexandria, VA

Pure Speed:
VO2 max:
Lactate threshold:
Endurance:

Datathletics

Runs | Power Curve | Profile | Blog | Log Out

Alex's Runs

- **Oct 4**: 6.4 miles in Alexandria, VA

- **Oct 3**: 3.1 miles in Washington, DC
- **Oct 1**: 9.6 miles in Occoquan, VA
- **Sept 29**: 8.0 miles in Alexandria, VA

Alex Pape
Alexandria, VA

Pure Speed:
VO2 max:
Lactate threshold:
Endurance:

Questions ?