X-Informatics Physics Use Case Part II Looking for Higgs Particle Events and Counting

July 6 2013

Geoffrey Fox

gcf@indiana.edu

http://www.infomall.org/X-InformaticsSpring2013/index.html

Associate Dean for Research, School of Informatics and Computing

Indiana University Bloomington

2013

Big Data Ecosystem in One Sentence

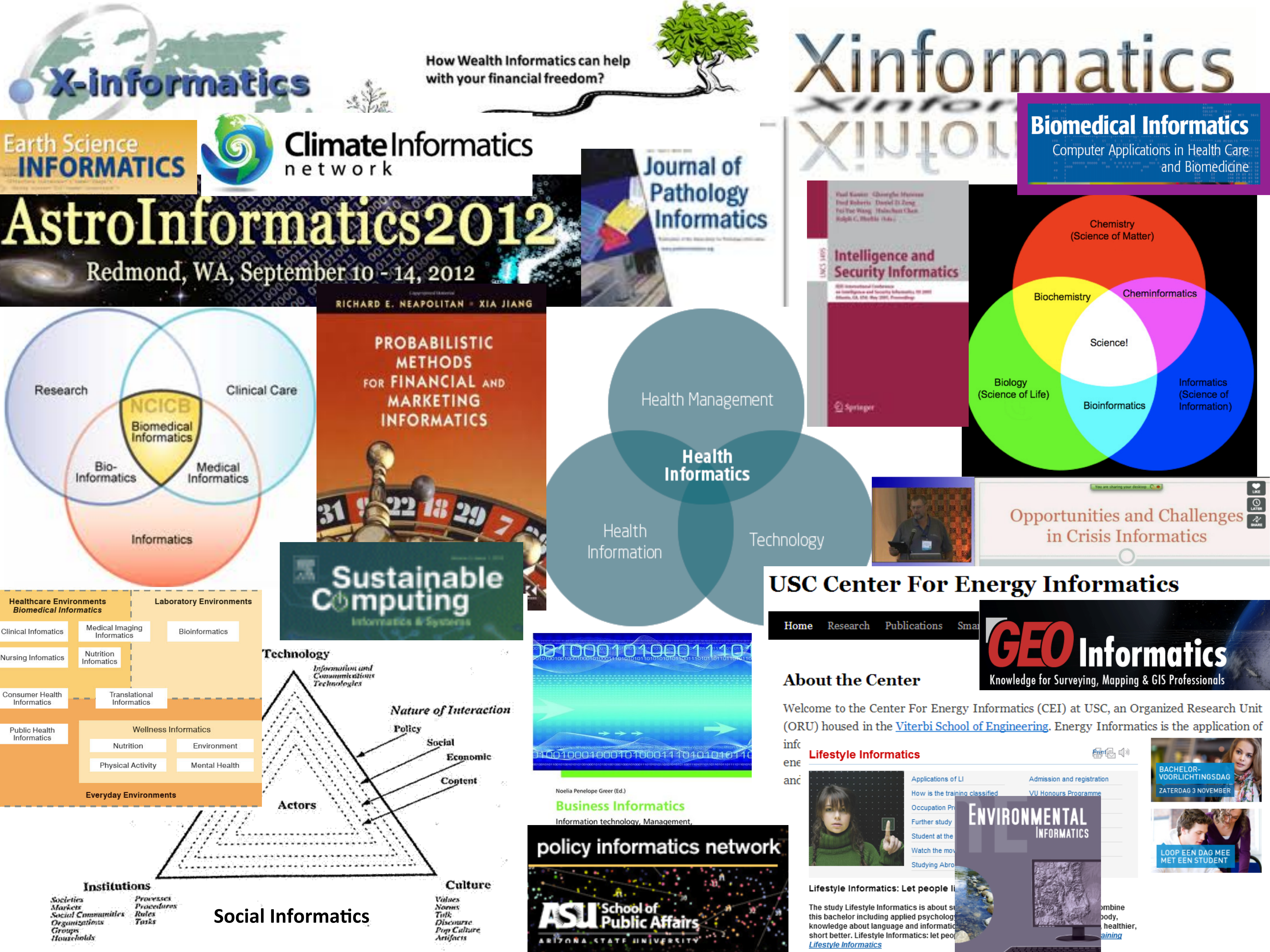
Use Clouds running Data Analytics Collaboratively processing Big Data to solve problems in X-Informatics (or e-X)

X = Astronomy, Biology, Biomedicine, Business, Chemistry, Climate, Crisis, Earth Science, Energy, Environment, Finance, Health, Intelligence, Lifestyle, Marketing, Medicine, Pathology, Policy, Radar, Security, Sensor, Social, Sustainability, Wealth and Wellness with more fields (physics) defined implicitly

Spans Industry and Science (research)

Education: Data Science see recent New York Times articles

http://datascience101.wordpress.com/2013/04/13/new-york-times-data-science-articles/


X-informatics
How Wealth Informatics can help with your financial freedom?
Xinformatics
Earth Science INFORMATICS
Climate Informatics network
Biomedical Informatics
Computer Applications in Health Care and Biomedicine
Journal of Pathology Informatics
AstroInformatics2012
Redmond, WA, September 10 - 14, 2012
Intelligence and Security Informatics
Chemistry (Science of Matter)
Biochemistry
Cheminformatics
Science!
Biology (Science of Life)
Bioinformatics
Informatics (Science of Information)
RICHARD E. NEAPOLITAN · XIA JIANG
PROBABILISTIC METHODS FOR FINANCIAL AND MARKETING INFORMATICS
Research
Clinical Care
NCICB
Biomedical Informatics
Bio-Informatics
Medical Informatics
Informatics
Health Management
Health Informatics
Health Information
Technology
Opportunities and Challenges in Crisis Informatics
Sustainable Computing
Informatics & Systems
USC Center For Energy Informatics
Home
Research
Publications
GEO Informatics
Knowledge for Surveying, Mapping & GIS Professionals
About the Center
Welcome to the Center For Energy Informatics (CEI) at USC, an Organized Research Unit (ORU) housed in the Viterbi School of Engineering. Energy Informatics is the application of
Healthcare Environments
Biomedical Informatics
Laboratory Environments
Clinical Infomatics
Medical Imaging Informatics
Bioinformatics
Nursing Infomatics
Nutrition Infomatics
Consumer Health Informatics
Translational Informatics
Public Health Informatics
Wellness Informatics
Nutrition
Environment
Physical Activity
Mental Health
Everyday Environments
Technology
Information and Communications Technologies
Nature of Interaction
Policy
Social
Economic
Content
Actors
Institutions
Societies
Markets
Social Communities
Organizations
Groups
Households
Processes
Procedures
Rules
Tasks
Culture
Values
Norms
Talk
Discourse
Pop Culture
Artifacts
Social Informatics
Noelia Penelope Greer (Ed.)
Business Informatics
Information technology, Management,
policy informatics network
ASU School of Public Affairs
ARIZONA STATE UNIVERSITY
Lifestyle Informatics
Applications of LI
How is the training classified
Admission and registration
VU Honours Programme
Further study
ENVIRONMENTAL INFORMATICS
BACHELOR-VOORLICHTINGSDAG
ZATERDAG 3 NOVEMBER
LOOP EEN DAG MEE MET EEN STUDENT
Lifestyle Informatics: Let people li
Lifestyle Informatics

Event Counting

Event Counting

- A lot of data analysis consists of setting up a process that gives "events" as a result
  - Take a survey of what people feel; event is voting of an individual
  - Build a software system and categorize each of your data; event is this categorization
  - Sensor Nets; event is result of a sensor measurement
- Often results are yes/no
  - Did person vote for Candidate X or not?
  - Did event fall into a certain bin of histogram or not?
  - Note event might have a result (cost of Hotel stay or Mass of Higgs) which is histogrammed; the decision as to which bin to go into is a yes/no decision

Generate a Physics Experiment with Python

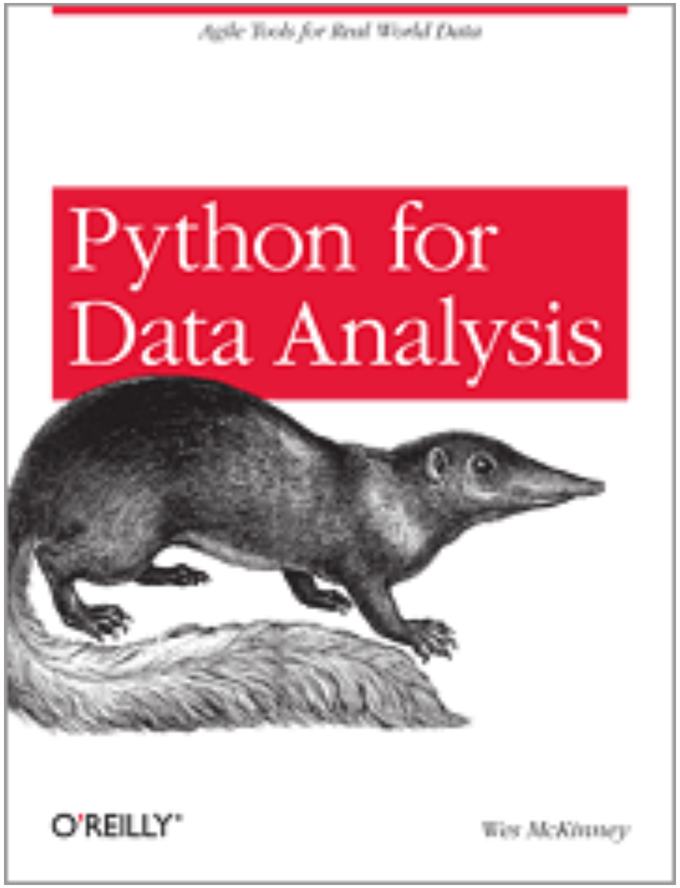
```
import numpy as np
import matplotlib.pyplot as plt
testrand = np.random.rand(42000)
Base = 110 + 30* np.random.rand(42000)

index = (1.0 - 0.5* (Base-110)/30) > testrand
gauss = 2 * np.random.randn(300) +126
Sloping = Base[index]
NarrowGauss = 0.5 * np.random.randn(300) +126
total = np.concatenate((Sloping, gauss))
NarrowTotal = np.concatenate((Sloping, NarrowGauss))
```

Python Resources

- http://www.enthought.com/products/epdgetstart.php
  Python distribution including NumPy and SciPy with plot package matplotlib
  - Also pandas and iPython

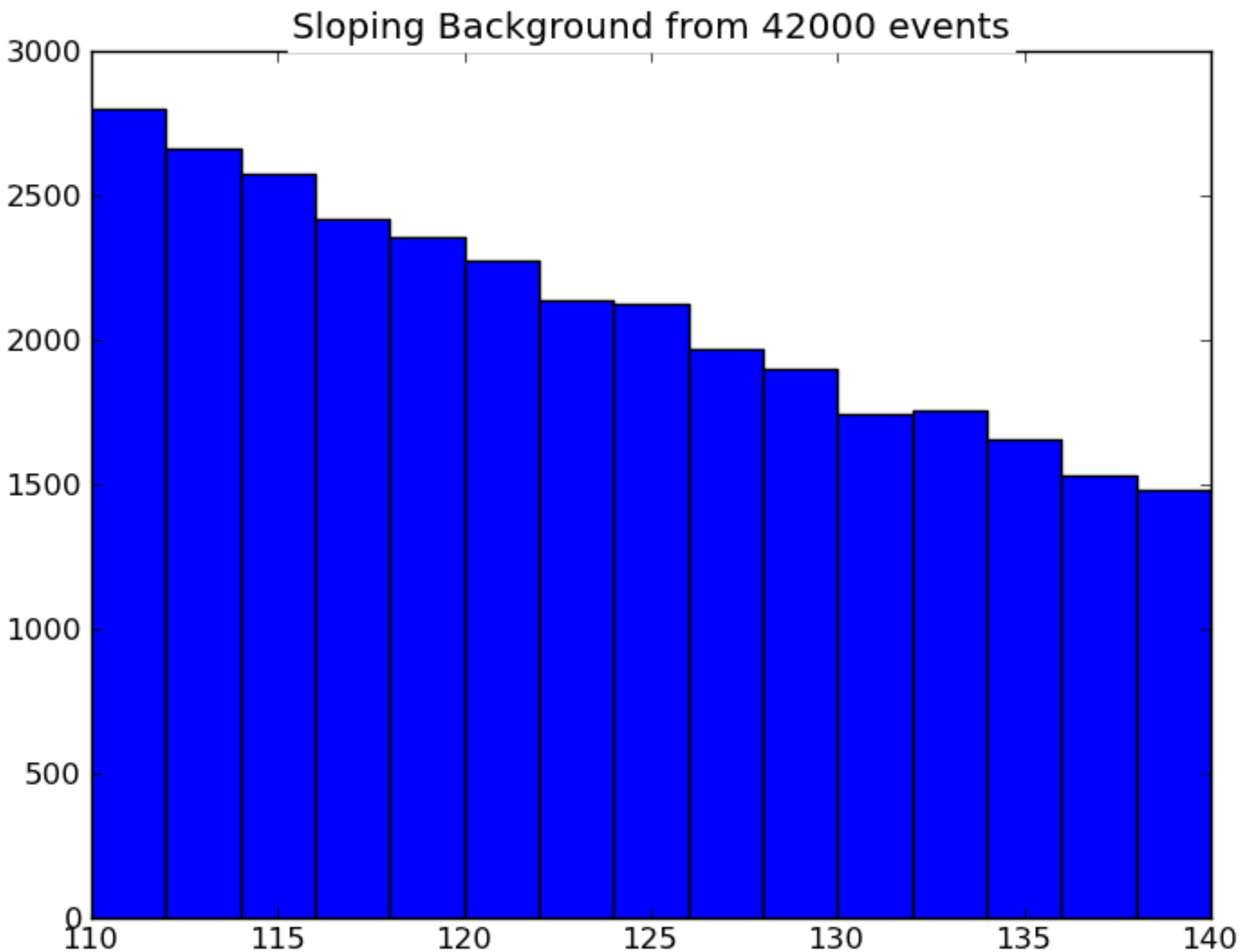
- **Python for Data Analysis**
  Agile Tools for Real World Data
  By [Wes McKinney]
- Publisher: O'Reilly Media
- Released: October 2012
- Pages: 472



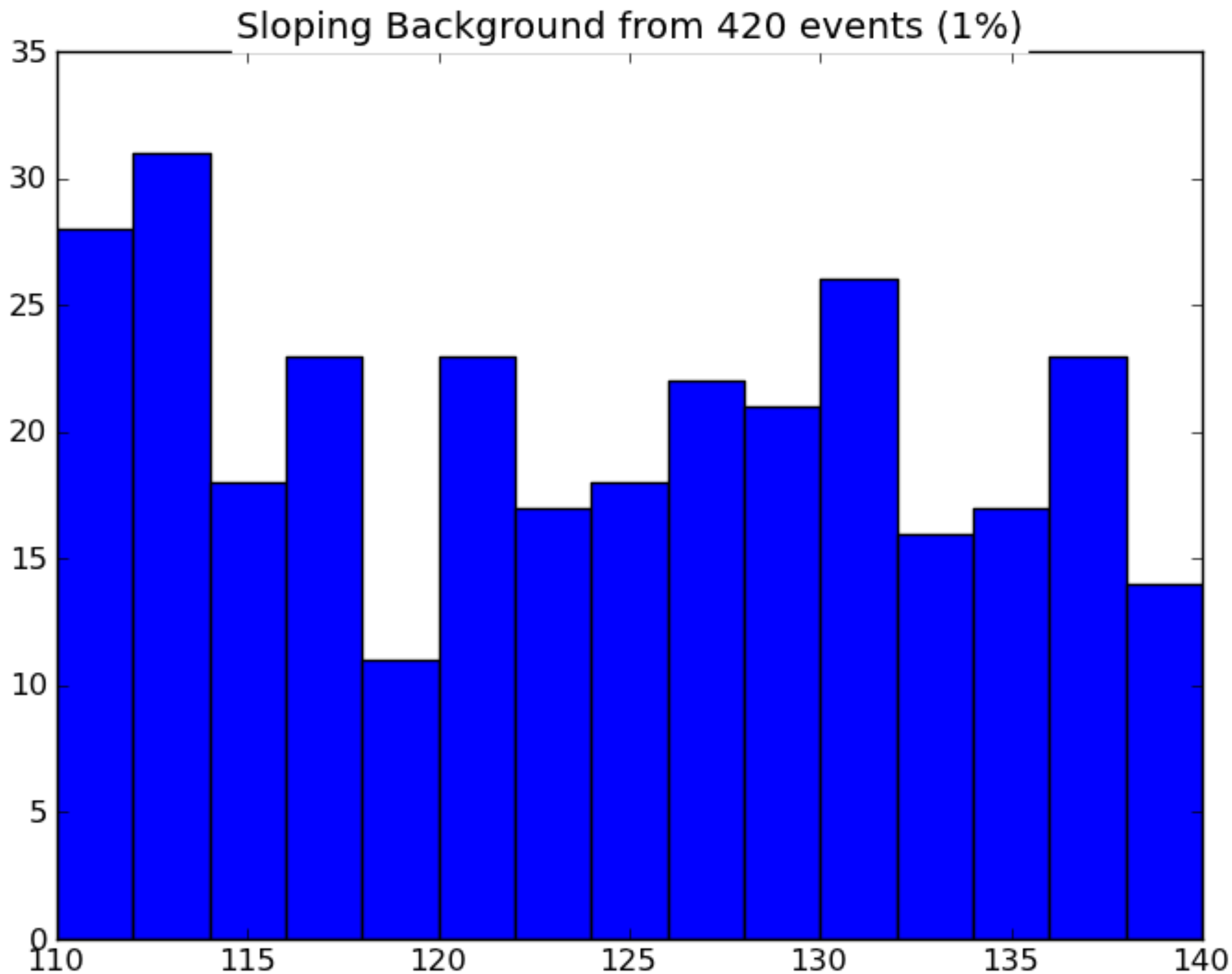
Examples of Event Counting

With Python examples of Signal plus Background

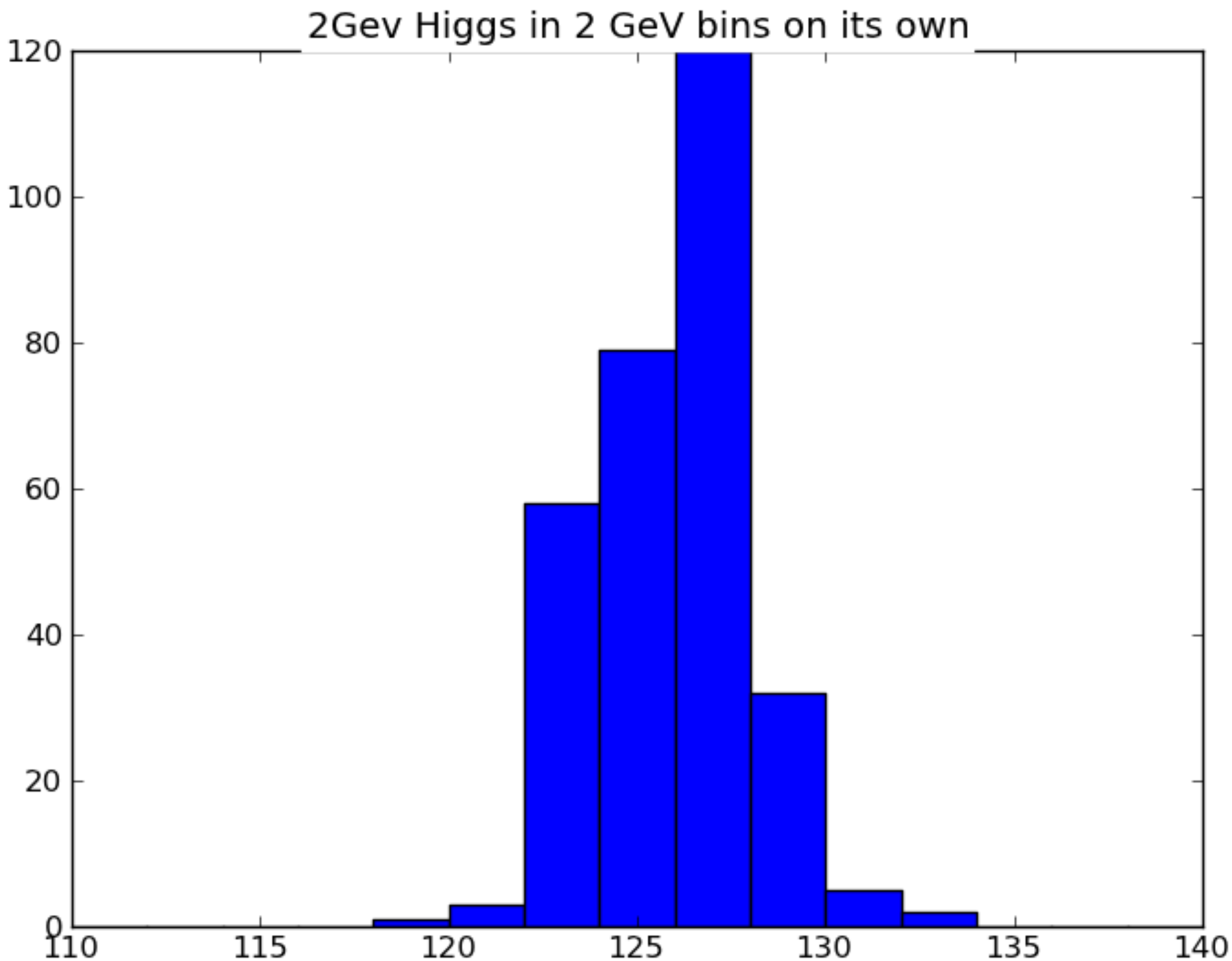
Sloping


Sloping Background from 42000 events
3000
2500
2000
1500
1000
500
0
110
115
120
125
130
135
140

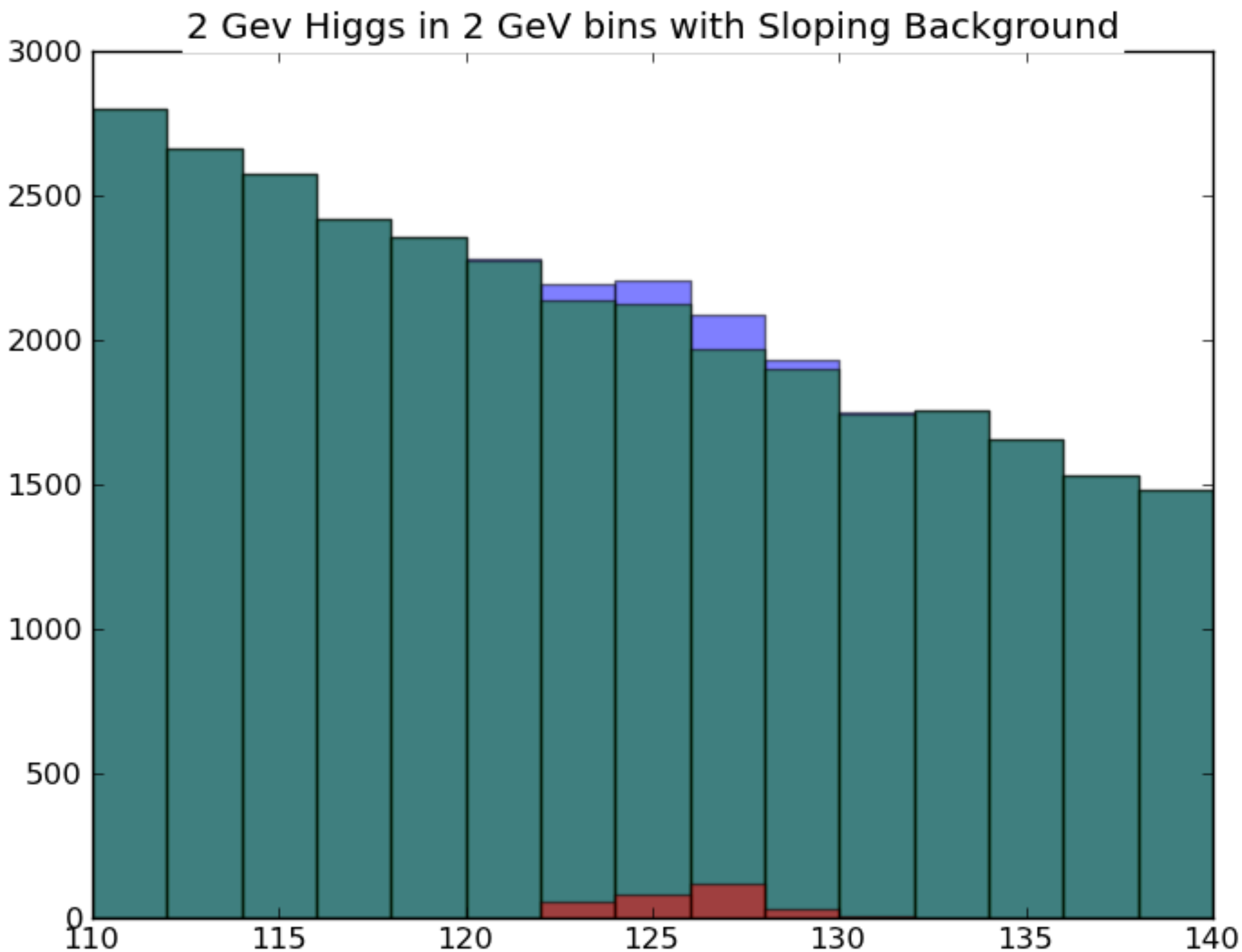
Sloping 1% Data



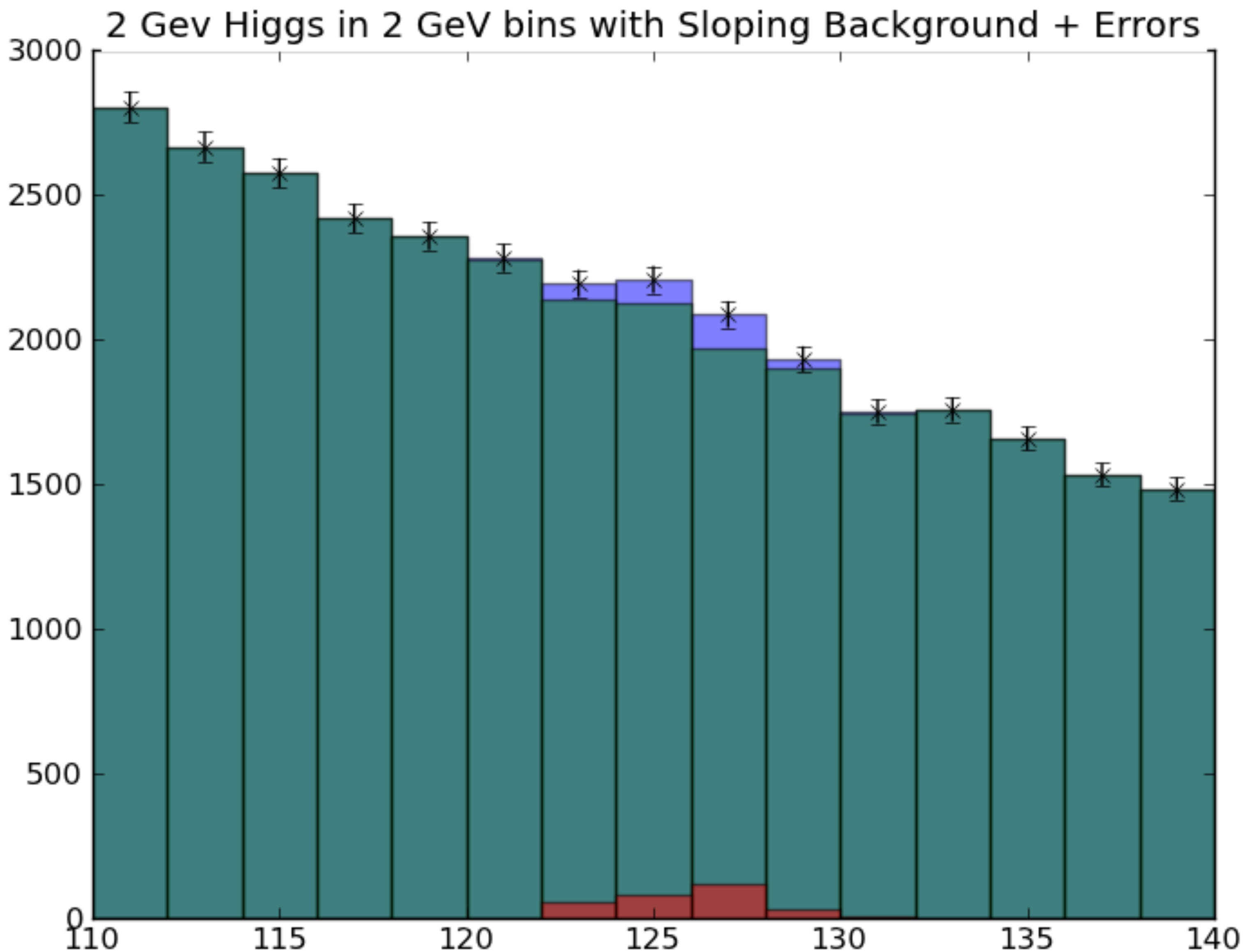
Higgs on its own



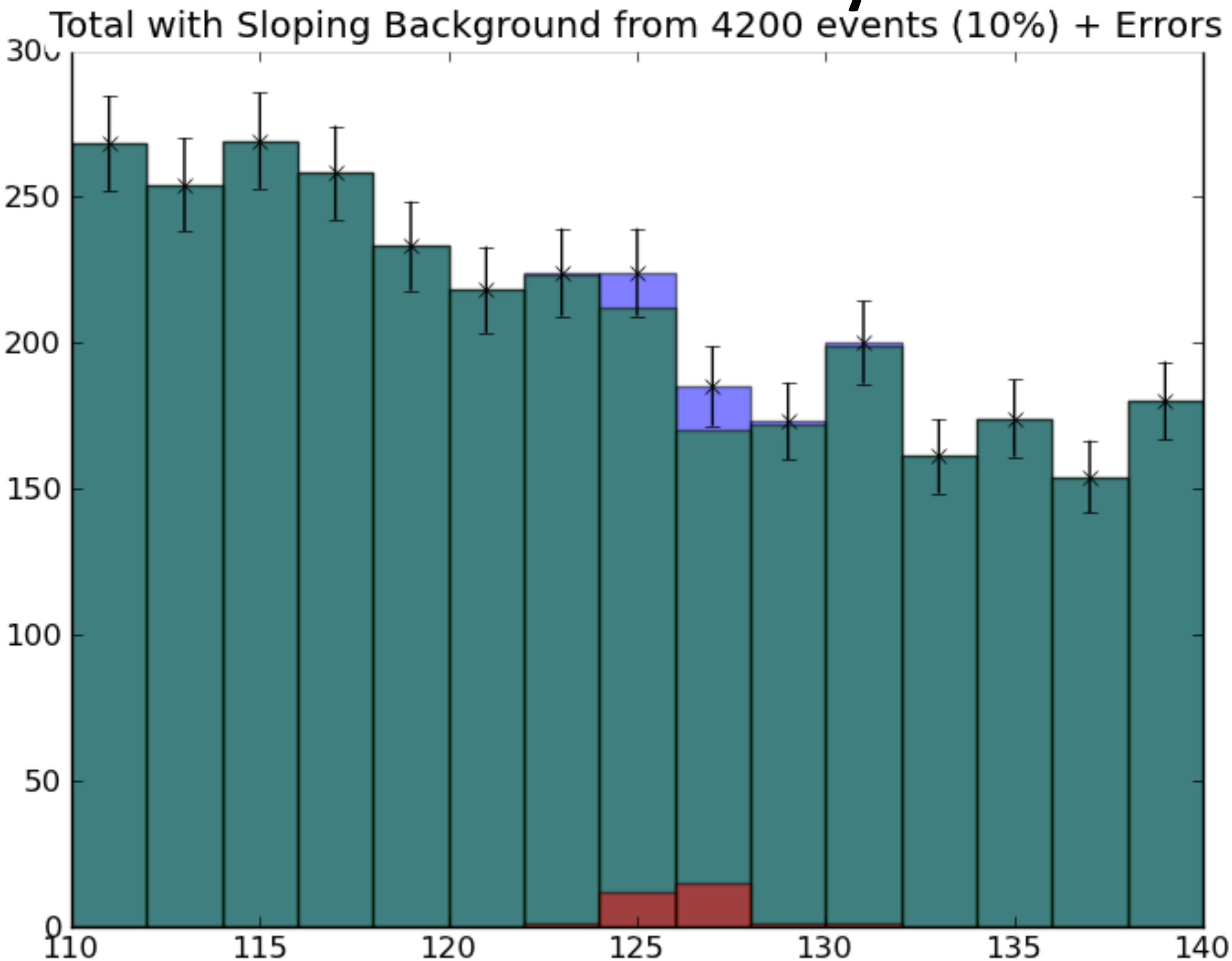
Actual Wide Higgs plus Sloping Background



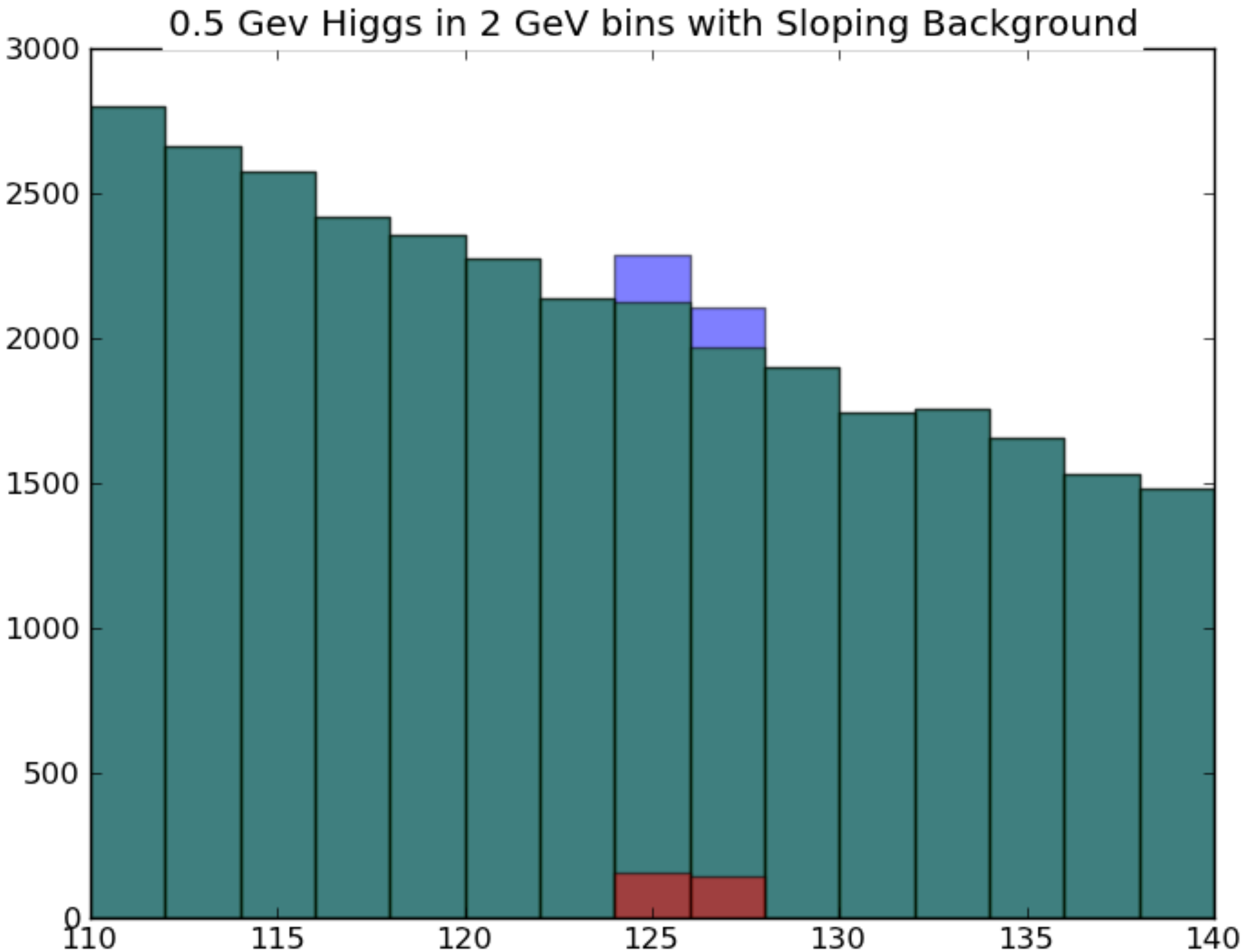
Actual Wide Higgs plus Sloping Background with errors



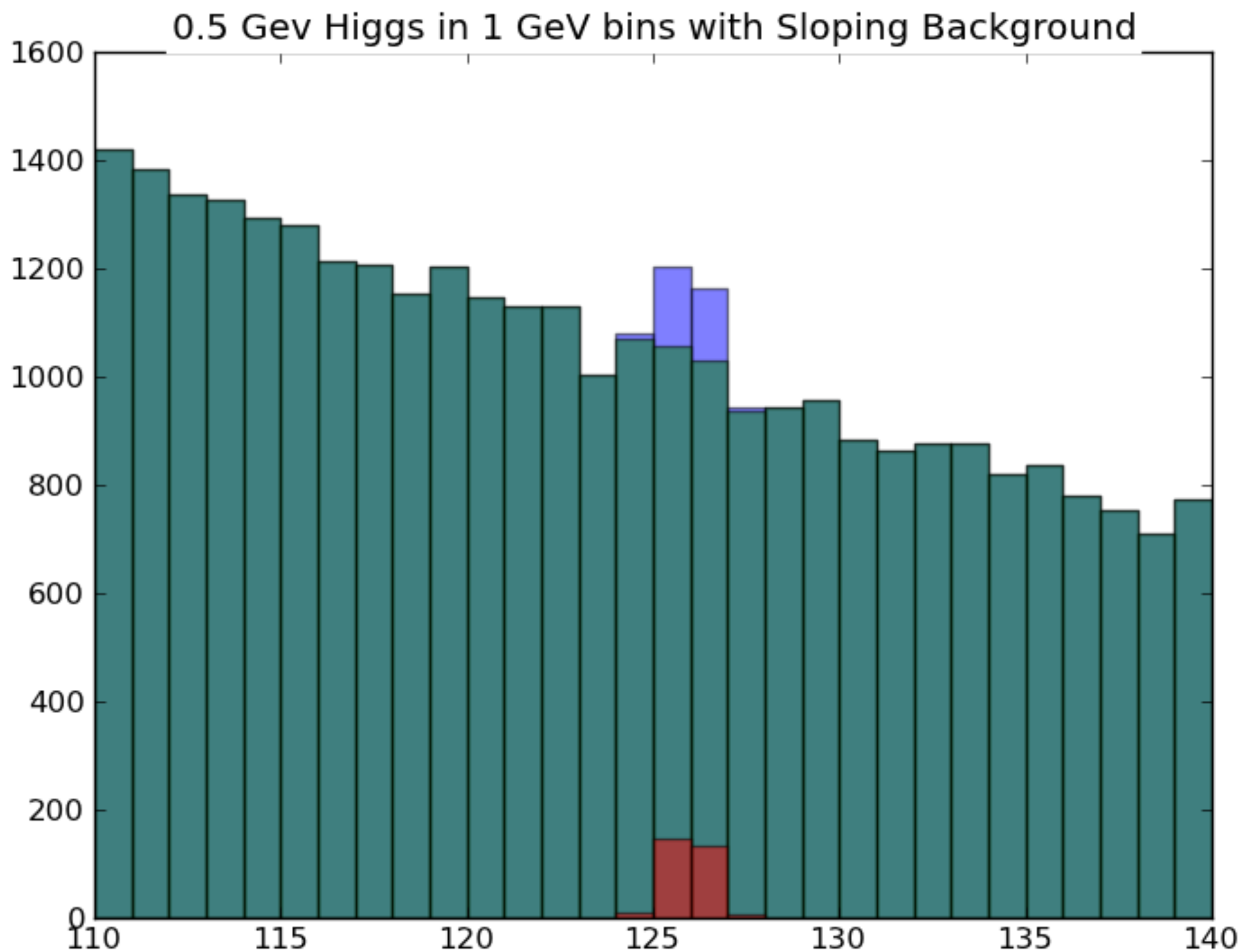
Actual Wide Higgs plus Sloping Background
Data Size divided by 10



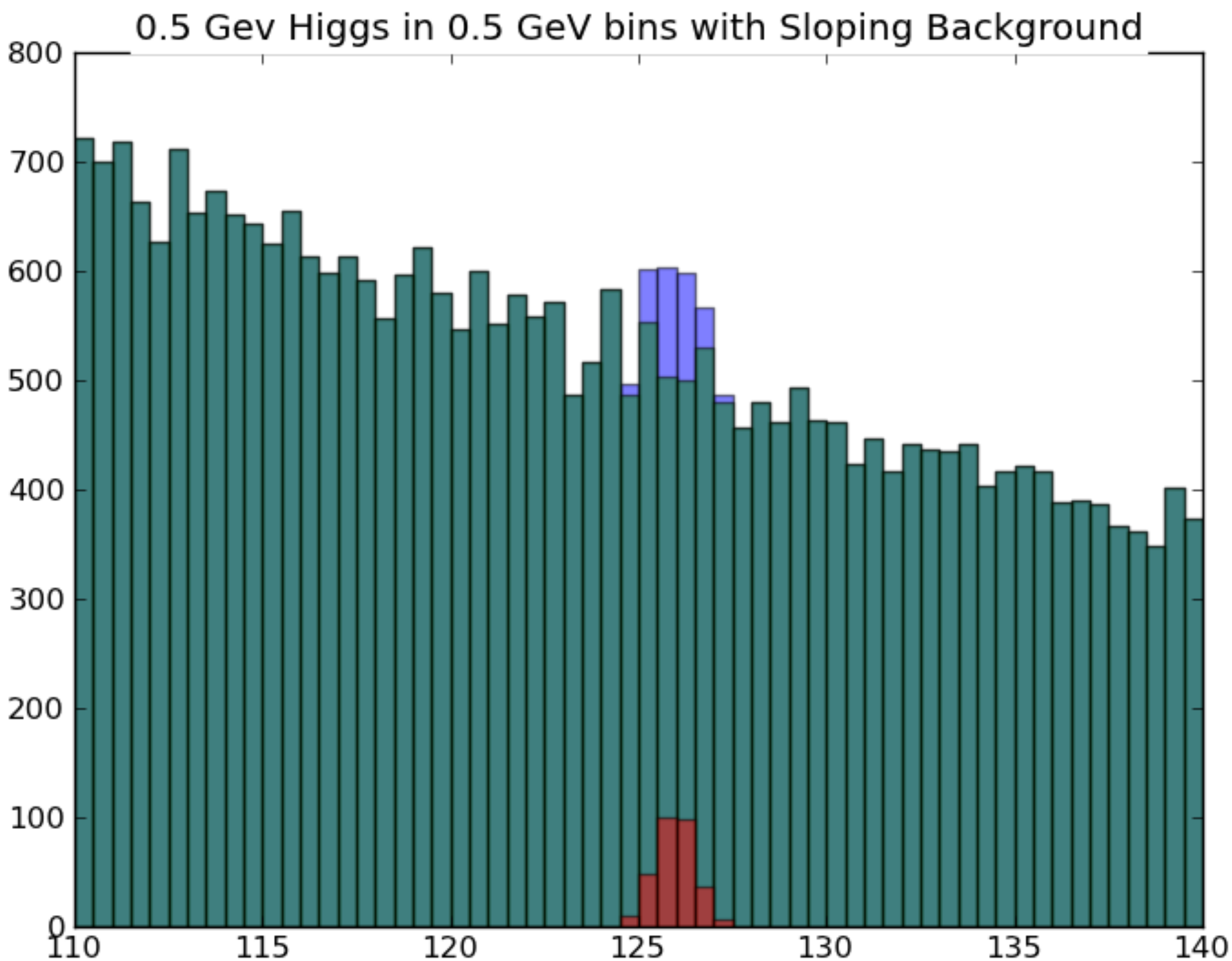
Narrow Higgs plus Sloping Background
2 GeV Bins



Narrow Higgs plus Sloping Background
1 GeV Bins



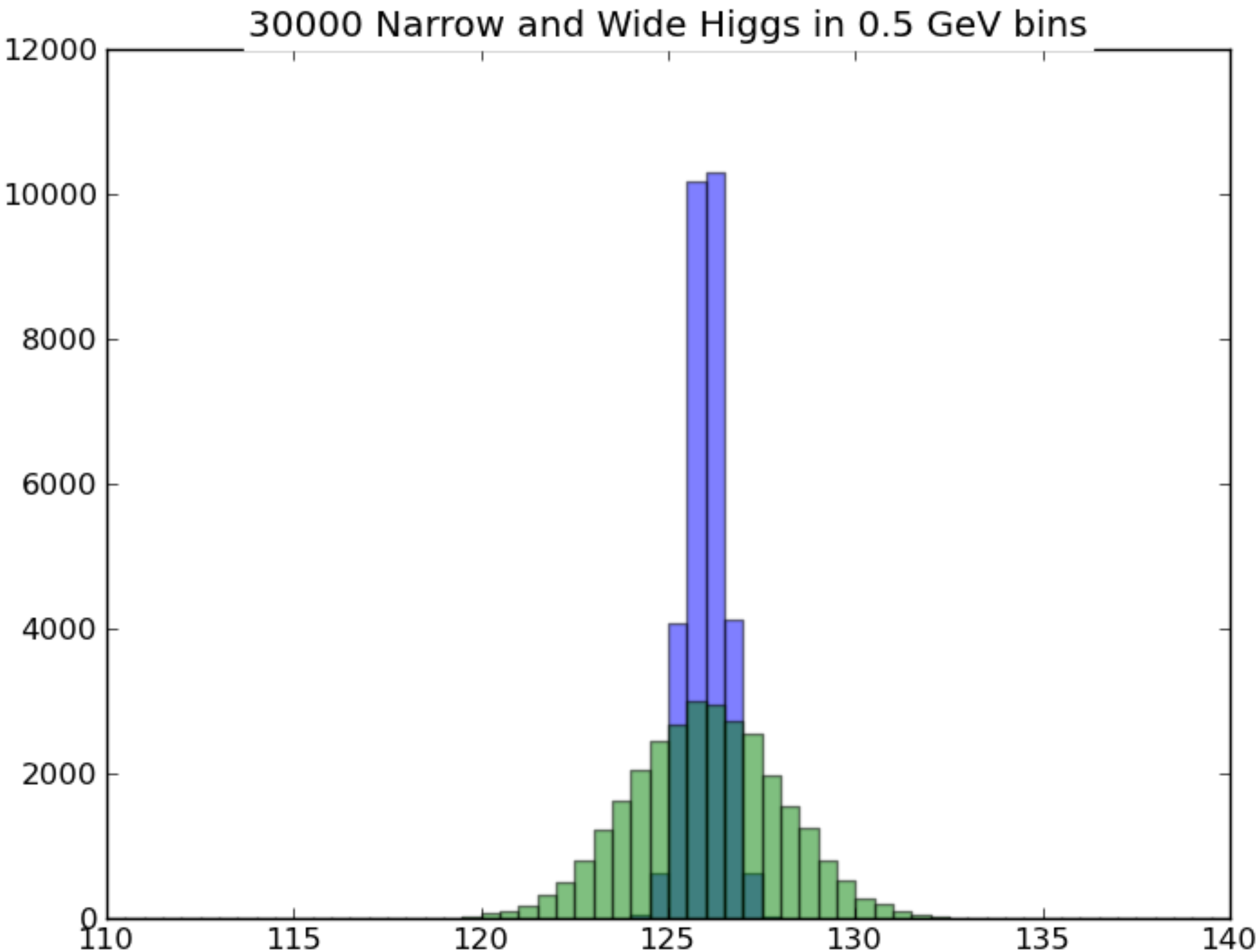
Narrow Higgs plus Sloping Background
0.5 GeV Bins



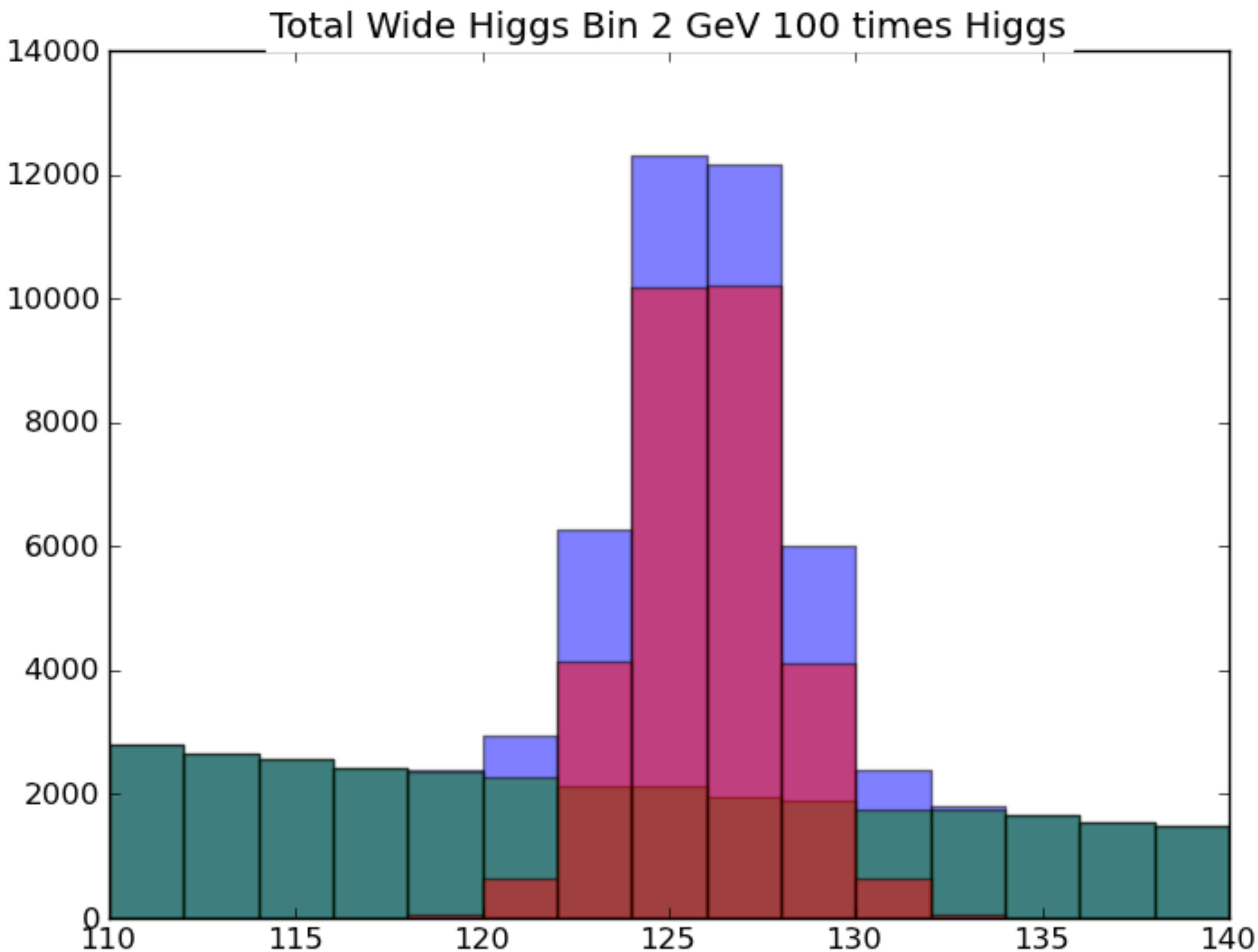
What Happens if more Higgs produced

- gaussbig = 2 * np.random.randn(30000) +126
- gaussnarrowbig = 0.5 * np.random.randn(30000) +126
- totalbig = np.concatenate((Sloping, gaussbig))

30,000 Higgs (Real Width and Narrow)



30,000 Higgs (Real Width) + Background



Simplification with flat background

- import numpy as np
- import matplotlib.pyplot as plt
- Base = 110 + 30* np.random.rand(42000)
- # Base is set of observations with an expected 2800 background events per bin
- # Note we assume here flat but in class I used a "sloping" curve that represented experiment better
- gauss = 2 * np.random.randn(300) +126
- # Gauss is Number of Higgs particles
- simpletotal = np.concatenate((Base, gauss))
- # simpletotal is Higgs+Background
- plt.figure("Total Wide Higgs Bin 2 GeV")
- values, binedges, junk = plt.hist(simpletotal, bins=15, range =(110,140), alpha = 0.5, color="green")
- centers = 0.5*(binedges[1:] + binedges[:-1])
- # centers is center of each bin
- # values is number of events in each bin
- # :-1 is same as :Largest Index-1
- # binedges[:-1] gets you lower limit of bin
- # 1: gives you array starts at second index (labelled 1 as first index 0)
- # binedges[1:] is upper limit of each bin
- # Note binedges has Number of Bins + 1 entries; centers has Number of Bins entries
- errors =sqrt(values)
- # errors is expected error for each bin
- plt.hist(Base, bins=15, range =(110,140), alpha = 0.5, color="blue")
- plt.hist(gauss, bins=15, range =(110,140), alpha = 1.0, color="red")
- plt.errorbar(centers, values, yerr = errors, ls='None', marker ='x', color = 'black', markersize= 6.0)

Uniform Background