Introduction

We will soon be conducting statistical analysis on the Higgs-Boson particle, the particles it decays into, and the jets that it reconstructs into. Prior to this investigation, we will explore the data that we will be using to conduct said investigation.
The purpose of this write-up is to become confortable with the data and concepts that will be used and referenced in the labs to come.

The topics that will be explored in this write up are as follows:

- The Standard Model
- The Higgs-Boson Particle
- The Large Hadron Collider
- The Data and Its Features
- Significant Data Characteristics

An Introduction to the Standard Model

Before proceeding, we must delcare and understand the following facts:

The Fundamental Forces

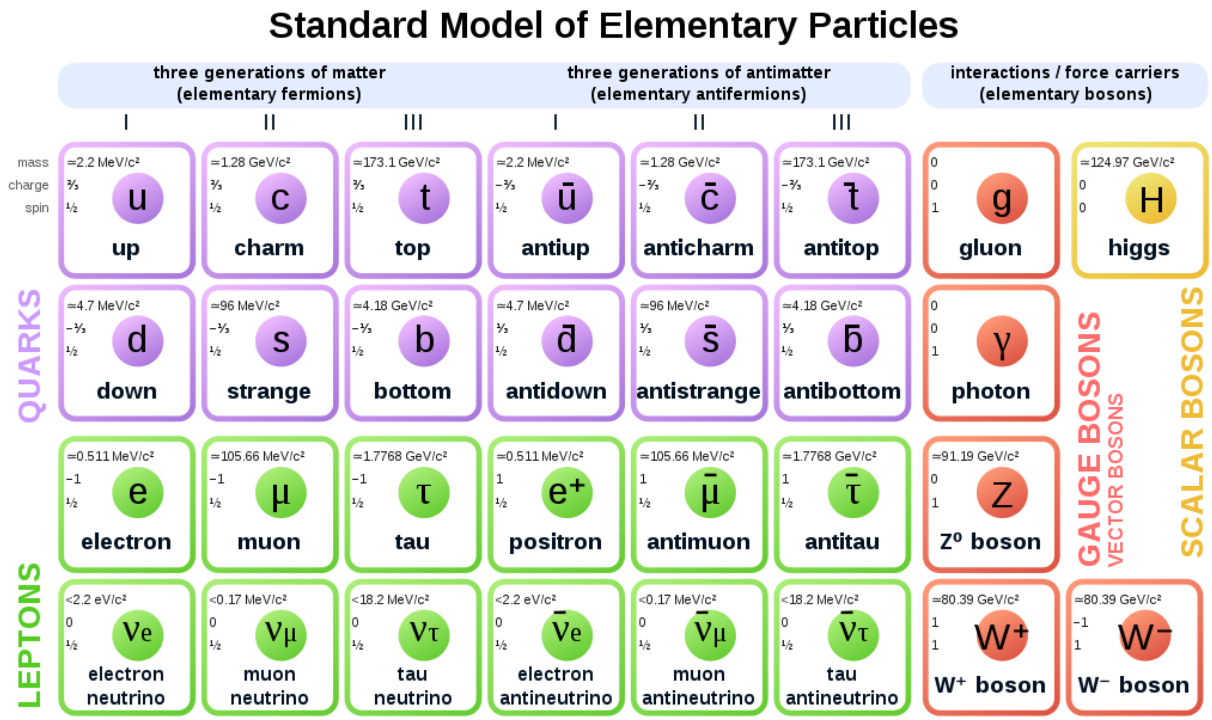
In the universe, there are four known fundamental forces:

- Electromagnetic
  - Explains how both moving and stationary particles interact
- Weak Interaction
  - The mechanism of interaction between subatomic particles that is responsible for the radioactive decay of atoms
- Strong Interactions
  - The attractive force that binds the elementary particles together
- Gravitational Force
  - The force that attracts any two objects wtith mass

Elementary Particles

In particle physics, there are things called "Elementary Particles". These are subatomic particles with no sub structure, therefore not composed of other particles.

The following is a visualization of the known Elementary Particles:



The Standard Model

With the above known, we can now explore the meaning of the "Standard Model": The standard model of particle physics is the theory describing three of the four known fundamental forces (excluding the gravitational force) in the universe, as well as classifying all known elementary particles. In other words, it is a way to explain the properties of forces between the elementary particles.

While the work to develop the standard model was conducted in stages throughout the 20th century, the current formulation was finalized in the mid-1970s upon experimental confirmation of the existence of quarks.

Since then, the confirmation of the top quark, tau neutrino, and the Higgs boson particles have added further credence to the Standard Model. In this write-up, we will be exploring the Higgs boson particle in detail.

The Higgs Boson Particle

As discussed, one of the particles included in the Standard Model is the boson, which the Higgs particle is classified as (hence the name).

A Brief History

This particle was named after Physicist Peter Higgs. In 1964 he attempted to explain why particles have math with the proposal of the Higgs mechanism. The Higgs mechanism, simply put, is the mechanism by which the Higgs field leads to the breaking of certain symmetry laws of the electroweak interaction. This field, by way of the Higgs mechanism, causes the gauge bosons of the weak force to be massive. The proposal of this Higgs mechanism implied the existence of the Higgs boson.

Its existence was finally confirmed in 2012 by ATLAS and CMS collaborations baed on collisions in the Large Hadron Collider (LHC) at CERN.

Known Characteristics of the Higgs Boson

The Higgs boson was difficult to detect due to the energy required to produce them and their vary rare production (1 in 10 billion at the LHC) even in the event of sufficient energy.

Some of the characteristics that will be of interest to us in the coming experiments are as follows:

- Mass = 125.18 +- .16 GeV/c
- Predicted Mean Lifetime = 1.56x10^-22 s
- Decays Into:
    - Bottom-antibottom pair (observed)
    - Two W bosons (observed)
    - Two gluons (predicted)
    - Two antitau pairs (observed)
    - Two Z bosons (observed)
    - Two photos (observed)
- Spin = 0

The Large Hadron Collider

What Problem the LHC Solves

Shortly after the 19th century declaration of the existence of atoms, physicist have been searching for the fundamental particles of natures.

Through the 1950s and 1960s, a high number of particles were found in collisions of particles from increasingly high-energy beams. These collisions and the need for even higher energy beams gave birth to the idea of what would become the world's largest machine: The Large Hadron Collider

How It Works

The shape of the LHC is extremely important to the success of the machine. The collider is contained in a 3.8 meter wide, concrete, circular tunnel with a circumference of 26.7 km.

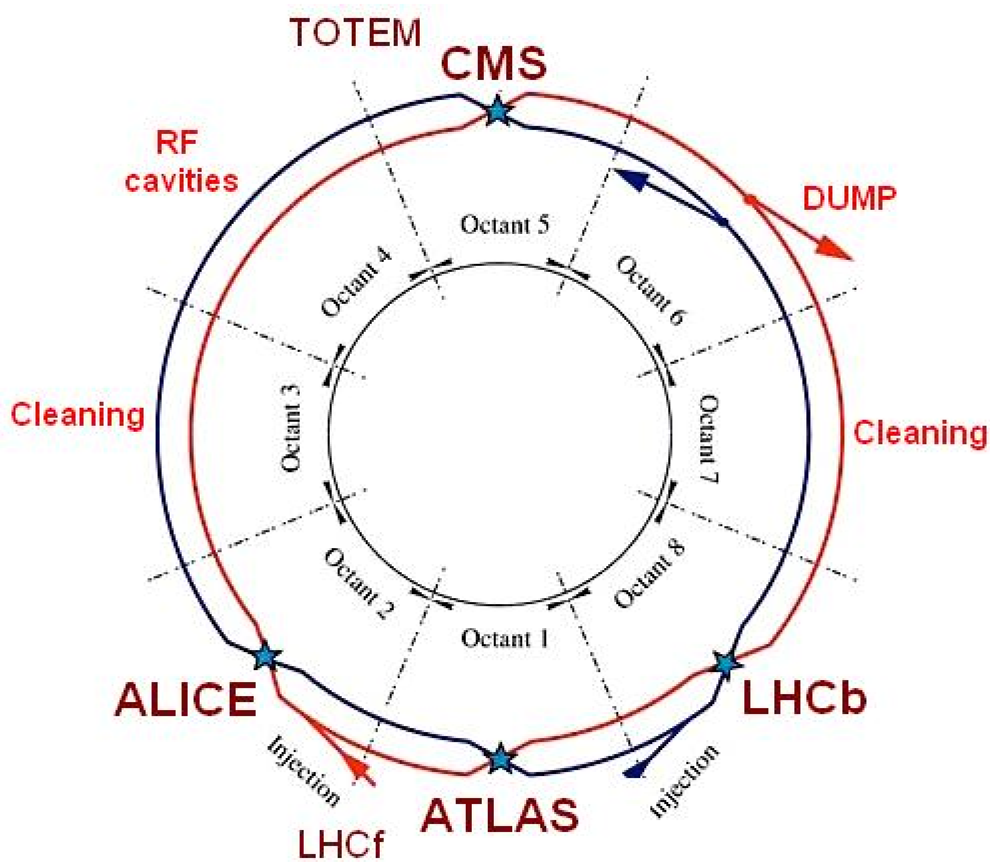
The collider tunnel contains two adjacent parralel beamlines, each containing a beam, which travel in opposite directions around the ring. These are the sources of the particles that are to be collided. These beams intersect at four points around the ring, which is where the actual collision takes place.

The particles would easily stray from a circular path were it not for over 1000 dipole magnets keeping the beams on that path. Even more magnets are needed, nearly 400 quadrupole magnets, to keep the beam focused (with stronger quadrupole magnets near intersections to maximize chances of interaction where the beams cross). In addition to those, there are many more magnets of higher pultipole order used to correct smaller imperfectiosn in the field geometry. All-in-all, there are about 9000 superconducting magnets in use to keep these beams on their circular path to collison.

Detectors

The LHC is only a piece of the machinery that works to detect particles. There are seven different detectors located around the LHC's intersection points. ATLAS and CMS, both paramount in the discovery of the Higgs boson, are both large general-purpose detectors. The others, LHCb, ALICE, TOTEM, MoEDAL, and LHCf are smaller than ATLAS and CMS and have more specific roles, or are used for very specialized research.

Below is a diagram of the layout at CERN:



Explanation of Data

The data we will be working with are simulated data, not actual data from the LHC. It has a specific transverse momentum range of 1000-12000 GeV, and includes two files which have the following data:

- QCD Background
  - Stands for "Quantum Chromodynamics"

- Generally refers to processes which lead to hadronic jets (narrow cone of hadrons and other particles produced by the formation of hadrons (hadronization) out of quarks and gluons)
- Higgs data (expected)
  - Expected outcome of QCD Background data after optimization and manipulation

Features of Data

There are 14 features included in each data set:

- **Transverse Momentum**
  - Defined by the equation $\sqrt{p_x^2 + p_y^2}$ for some
  - This is the component momentum perpendicular to the beam line

- **Pseudorapidity**
  - Defined by the equation η = -ln*$\tan(\frac{\theta}{2})$ $\frac{-1}{2}$ln $\frac{|p|+p_L}{|p|-p_L}$
  - Spacial coordinate describing the angle of a particle relative to the beam axis

- **Azimuthal Angle**
  - Defined by the equation $\cos^{-1}(\frac{x}{r})$ for some plane x, and some vector r
  - The angle between some plane perpendicular to the beam, and a vector

- **Invariant Mass**
  - Defined by the equation $E^2=p^2+m^2$ for some mass m and some momentum p
  - Mass of particle

- **2-Point E_{CF} Ratio**
  - Defined by the equation
  $$\sum_{i<j\in J} p_{T,i}\, p_{T,j}\, \Delta R_{ij} * \frac{1}{p_{T,J}^2}$$
  - 2-Point energy correlation function for quark/gluon discrimination

- **3-Point E_{CF} Ratio**
  - Defined by the equation
  $$\sum_{i<j<k\in J} p_{T,i}\, p_{T,j}\, p_{T,k}\, \Delta R_{ij}\, \Delta R_{ik}\, \Delta R_{jk} * \frac{1}{p_{T,J}^3}$$
  - 3-Point energy correlation function for W/Z/Higgs boson identification

- **3-to-2 Point E_{CF} Ratio (D_2)**
  - Defined by the equation $E_{CF3}\frac{E_{CF1}}{E_{CF2}}^3$
  - Ratio of energy correlation functions E_{CF2} and E_{CF3}

- **Angularity**
  - Defined by the equation
  $$\frac{1}{m_J}\sum_{i\in J} E_i \sin^{-2}(\Theta_i) * \cos^3(\Theta_i)$$

- Is suggested as a separator between QCD and heavy-object jets

- **N-Subjettiness**
  - Defined by the equation
$$\tau_N = \frac{1}{d_0} \sum_{k=1}^{M} (p_{T,k} \times \Delta R_{min,k})$$
  - A jet shape designed to identify boosted hadronically-decaying (decaying and producing a hadron) objects

- **KtDeltaR**
  - Defined by the equation ΔR=$\sqrt{\Delta\eta^2 + \Delta\varphi^2}$
  - Jet substructure moment

Exploring the Data

Setup:

In [2]:

```
import pickle
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats,signal

higg = open ("higgs_100000_pt_1000_1200.pkl",'rb')
higgD = pickle.load(higg)

qcd = open ("qcd_100000_pt_1000_1200.pkl",'rb')
qcdD = pickle.load(qcd)

higgsPt=higgD['pt']
higgsEta=higgD['eta']
higgsPhi=higgD['phi']
higgsMass=higgD['mass']
higgsEe2=higgD['ee2']
higgsEe3=higgD['ee3']
higgsD2=higgD['d2']
higgsAng=higgD['angularity']
higgsT1=higgD['t1']
higgsT2=higgD['t2']
higgsT3=higgD['t3']
higgsT21=higgD['t21']
higgsT32=higgD['t32']
higgsKtD=higgD['KtDeltaR']
qcdPt=qcdD['pt']
qcdEta=qcdD['eta']
qcdPhi=qcdD['phi']
qcdMass=qcdD['mass']
qcdEe2=qcdD['ee2']
qcdEe3=qcdD['ee3']
qcdD2=qcdD['d2']
qcdAng=qcdD['angularity']
qcdT1=qcdD['t1']
qcdT2=qcdD['t2']
qcdT3=qcdD['t3']
qcdT21=qcdD['t21']
qcdT32=qcdD['t32']
qcdKtD=qcdD['KtDeltaR']
```

Mass

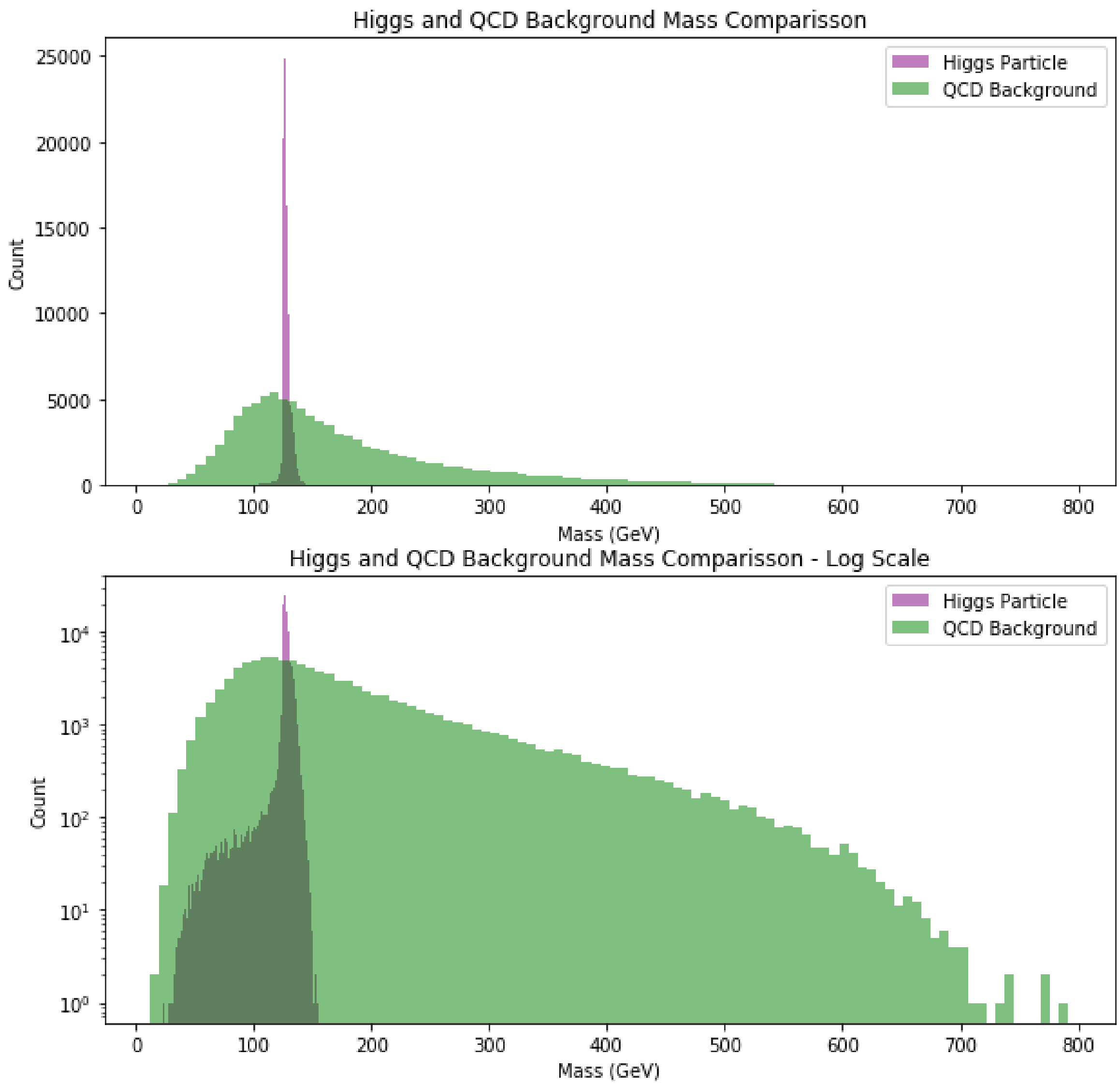
Below I have graphed the mass from both the Higgs data and the QCD background data.

You can see on both the linear and log-scale graphs that the Higgs data has a clear spike at approximately 125.

Given that this is a visualization of the mass, one can conclude that, given some data where a Higgs boson particle is expected, the mass of that particle would be approximately 125 GeV.

As previously discussed, the Higgs boson does have a mass of 125 GeV, and such a deduction of this visualization would be accurate.

In [83]:
```
fig, ax = plt.subplots(2, 1, figsize=(10,10))
ax[0].hist(higgsMass,bins=100,color='purple',alpha=0.5, label='Higgs Particle
ax[0].hist(qcdMass,bins=100,color='green',alpha=0.5, label='QCD Background')
ax[1].hist(higgsMass,bins=100,color='purple',alpha=0.5, label='Higgs Particle
ax[1].hist(qcdMass,bins=100,color='green',alpha=0.5, label='QCD Background')
ax[1].set_yscale('log')
ax[0].set_title("Higgs and QCD Background Mass Comparisson")
ax[0].legend(loc='upper right')
ax[1].legend(loc='upper right')
ax[1].set_title("Higgs and QCD Background Mass Comparisson - Log Scale")
ax[0].set_ylabel("Count")
ax[0].set_xlabel("Mass (GeV)")
ax[1].set_ylabel("Count")
ax[1].set_xlabel("Mass (GeV)")
plt.show()
```



Transverse Momentum

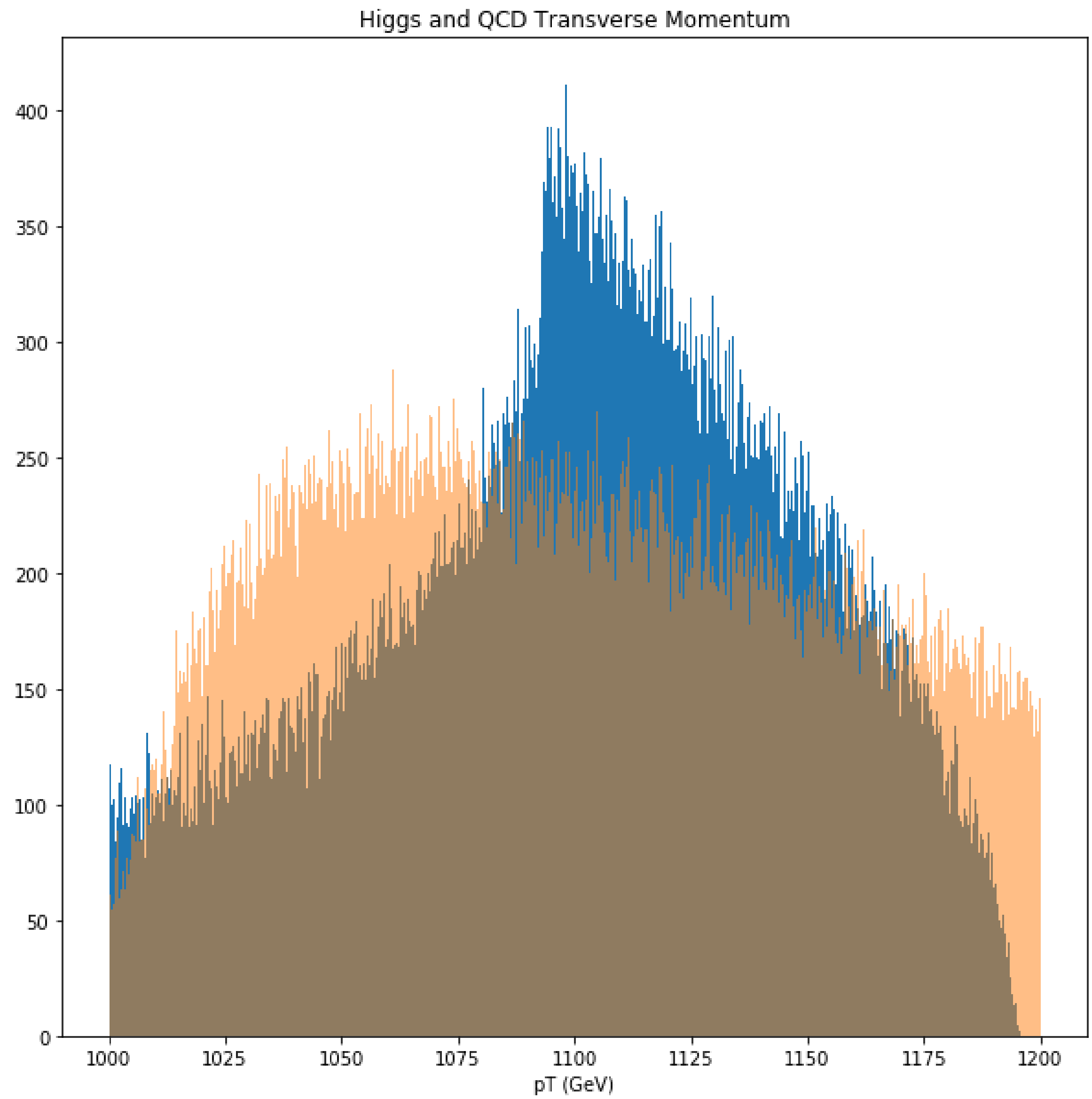
Shown below are the histograms of pT for both the background and higgs data. They will again be graphed on the same visualization in order to truly appreciate some of the features.

The QCD background in light orange, and the Higgs data in blue, we can see an interesting feature approaching the 1200 GeV pT limit (as defined by our data range).

Note that the background is clearly continuing, and cut off, when the data range ends. However, the Higgs data we see that it rapidly tapers off just before 1200 GeV at an estimated range of 1190-1199 GeV. At this time in the lab procedure, the significance (if any) of the transverse momentum dropping off here is unknown.

Another interesting feature that may prove useful in the future experiment is the dramatic increase of the Higgs data at just below 1100 GeV. The QCD background shows no such jump at this point. Again, at this time in the exploration of the data, the significance of this point is unknown (if there is any significance).

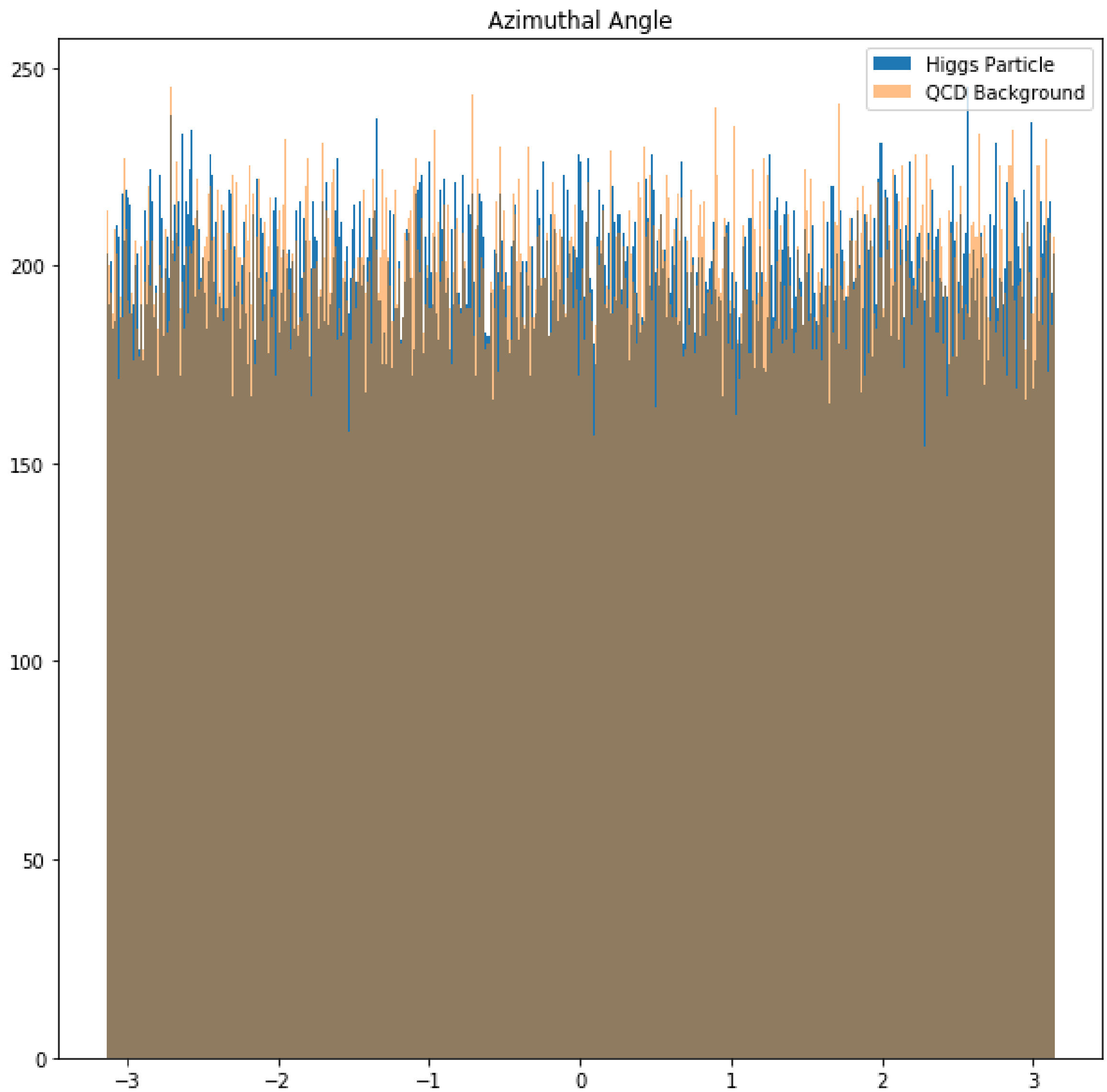
In [8]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsPt,500)
ax.hist(qcdPt,500,alpha=0.5)
ax.set_title("Higgs and QCD Transverse Momentum")
ax.set_xlabel("pT (GeV)")

plt.show()
```



Azimuthal Angle

There are no distinguishing features between the QCD background and the Higgs data for the Azimuthal angle. While the data itself may be useful, there is nothing in this feature that we could use to optimize the QCD background.

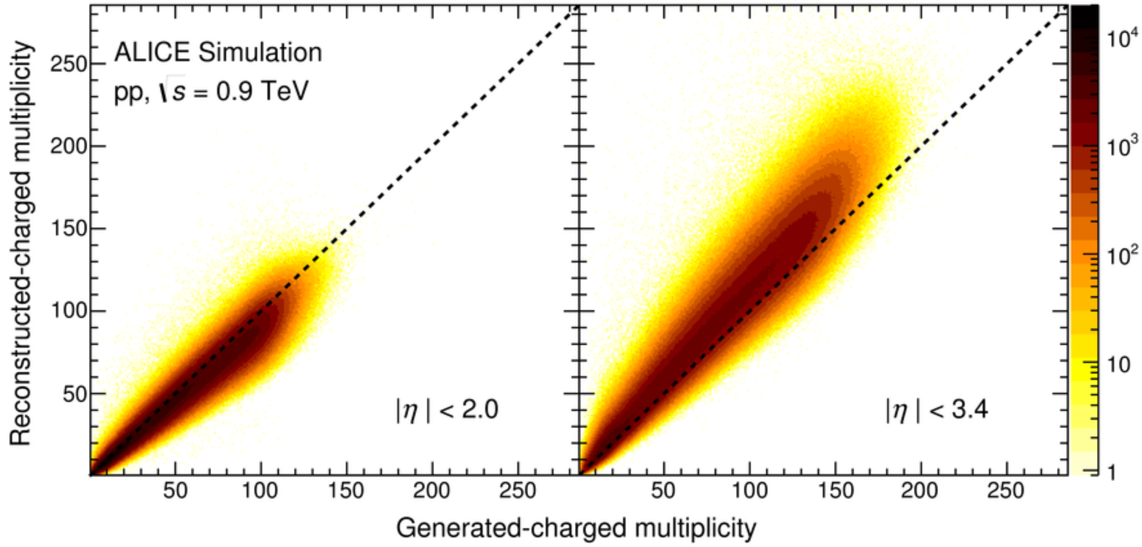
In [12]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsPhi,500,label='Higgs Particle')
ax.hist(qcdPhi,500,alpha=.5,label='QCD Background')
ax.set_title("Azimuthal Angle")
ax.legend(loc='upper right')
plt.show()
```



Pseudorapidity

In particle physics, an angle of zero is usually lalong the beam axis. This means that particles with high pseudorapidity values are generally lost.

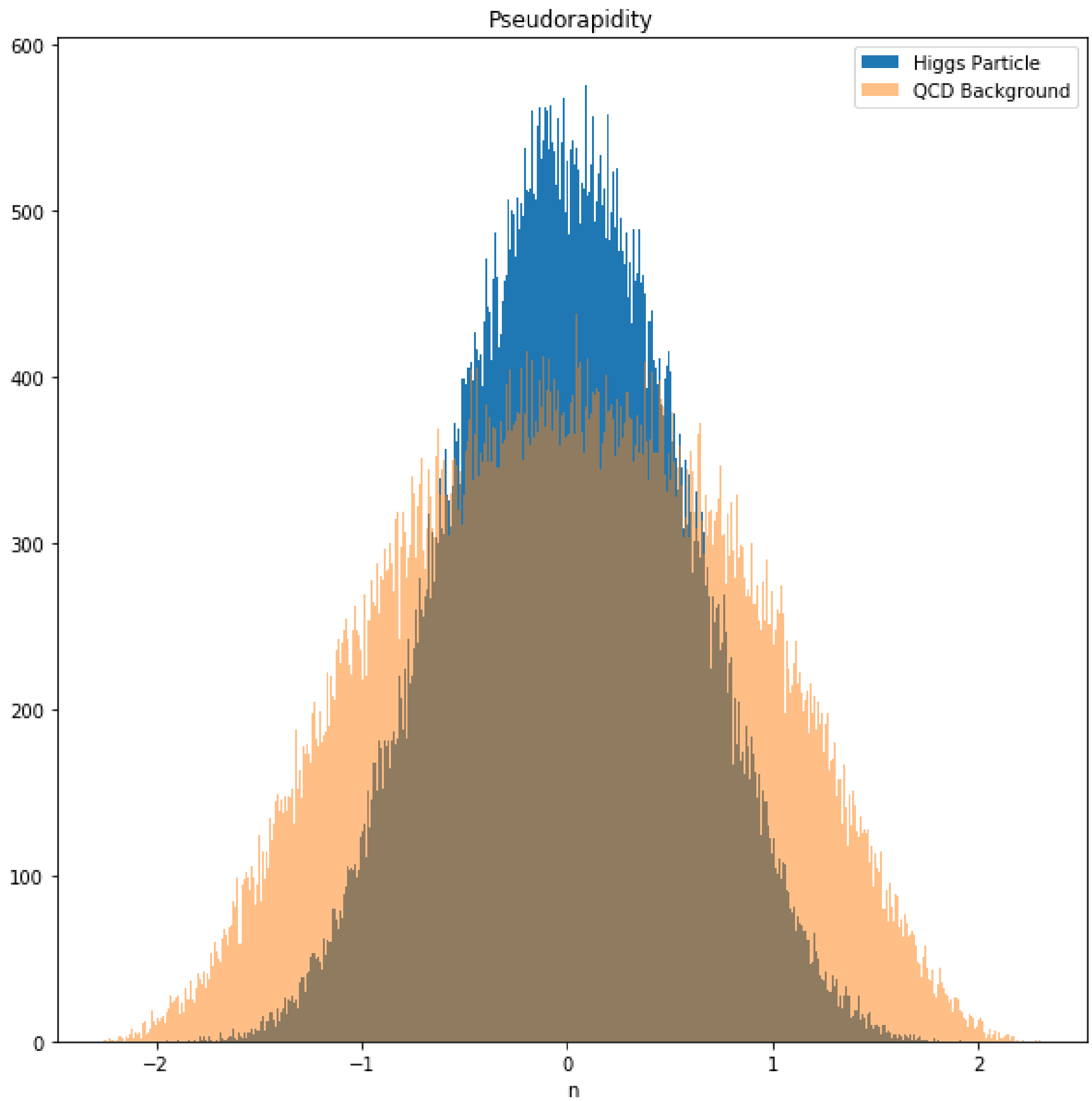
Here is a nice visualization of this concept:



With that in mind, we can comfortably choose to leave out much of the data (possibly even outside of a 2σ signal) to reduce the number of background measurements we have.

With the below histogram, we can confirm that the Higgs data does indeed have a mean of 0.

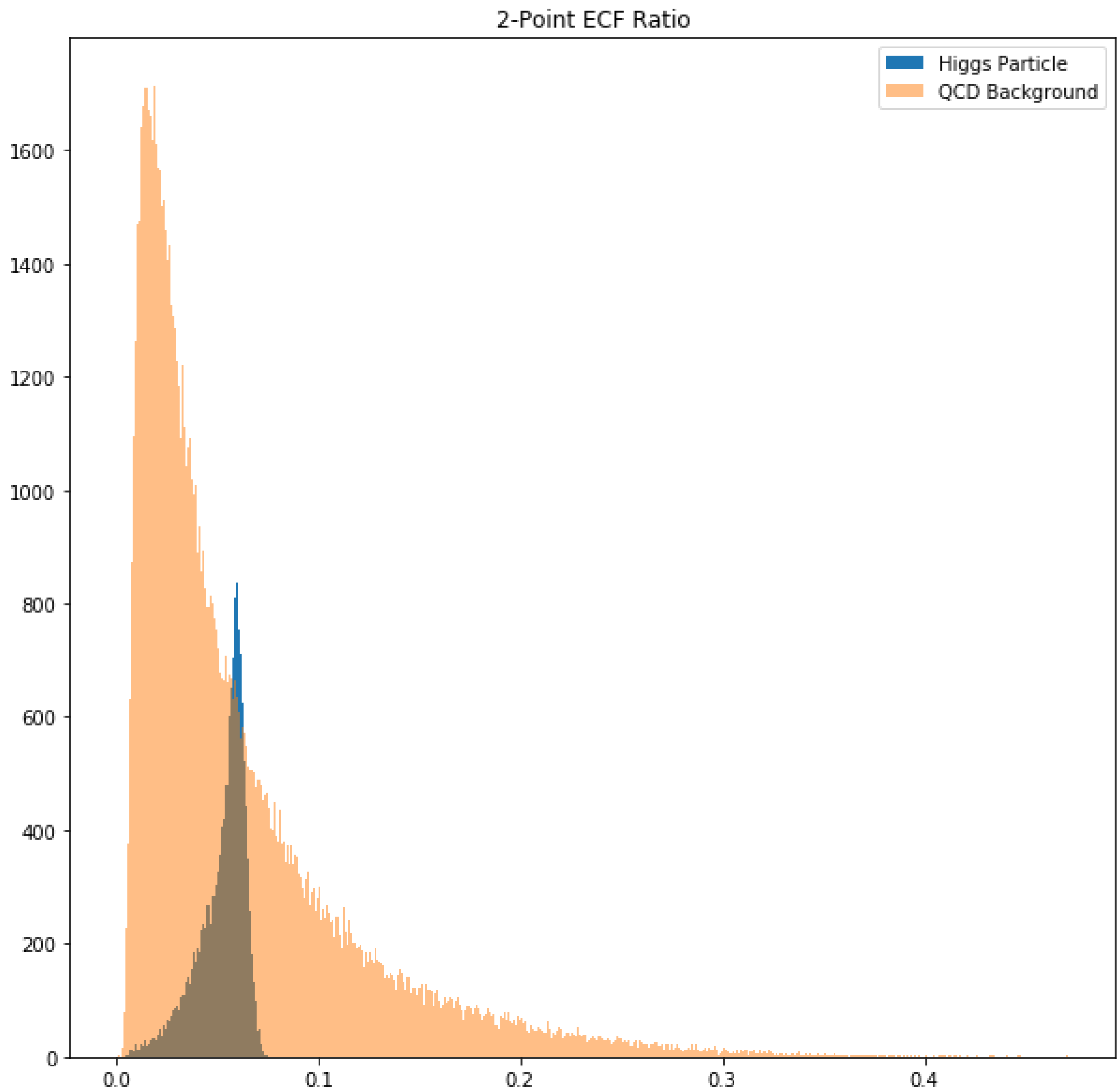
In [32]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsEta,500,label='Higgs Particle')
ax.hist(qcdEta,500,alpha=.5,label='QCD Background')
ax.set_title("Pseudorapidity")
ax.set_xlabel("n")
ax.legend(loc='upper right')
plt.show()
```



2-Point E_{CF} Ratio

I suspect, given that I have read that this is used in Gluon/Quark discrimination, that the peak of the Higgs particle data is us detecting Gluons and Quarks that have been produced from the beams colliding.

In [14]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsEe2,500,label='Higgs Particle')
ax.hist(qcdEe2,500,alpha=.5,label='QCD Background')
ax.set_title("2-Point ECF Ratio")
ax.legend(loc='upper right')
plt.show()
```



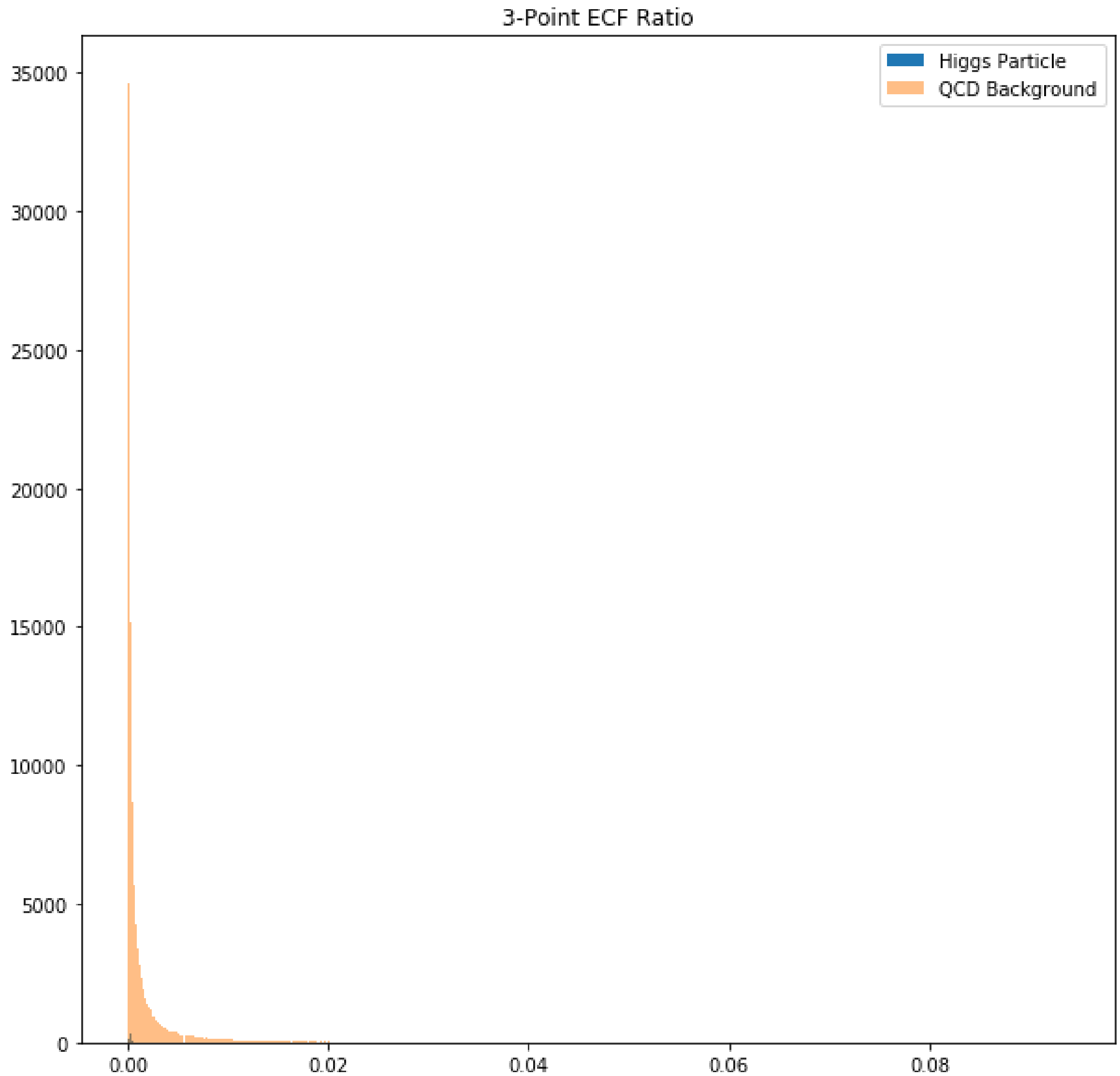
3-Point E_{CF} Ratio

The Higgs data is essentially non-existent in this feature as compared to the background. I would have expected to see a better representation for the Higgs particle here given that this data should represent a function for boson identification.

However, the fact that the Higgs data is difficult to see in comparison to the QCD background could demonstrate the rarity of actually observing a Higgs boson produced from beam collisions.

In [15]:

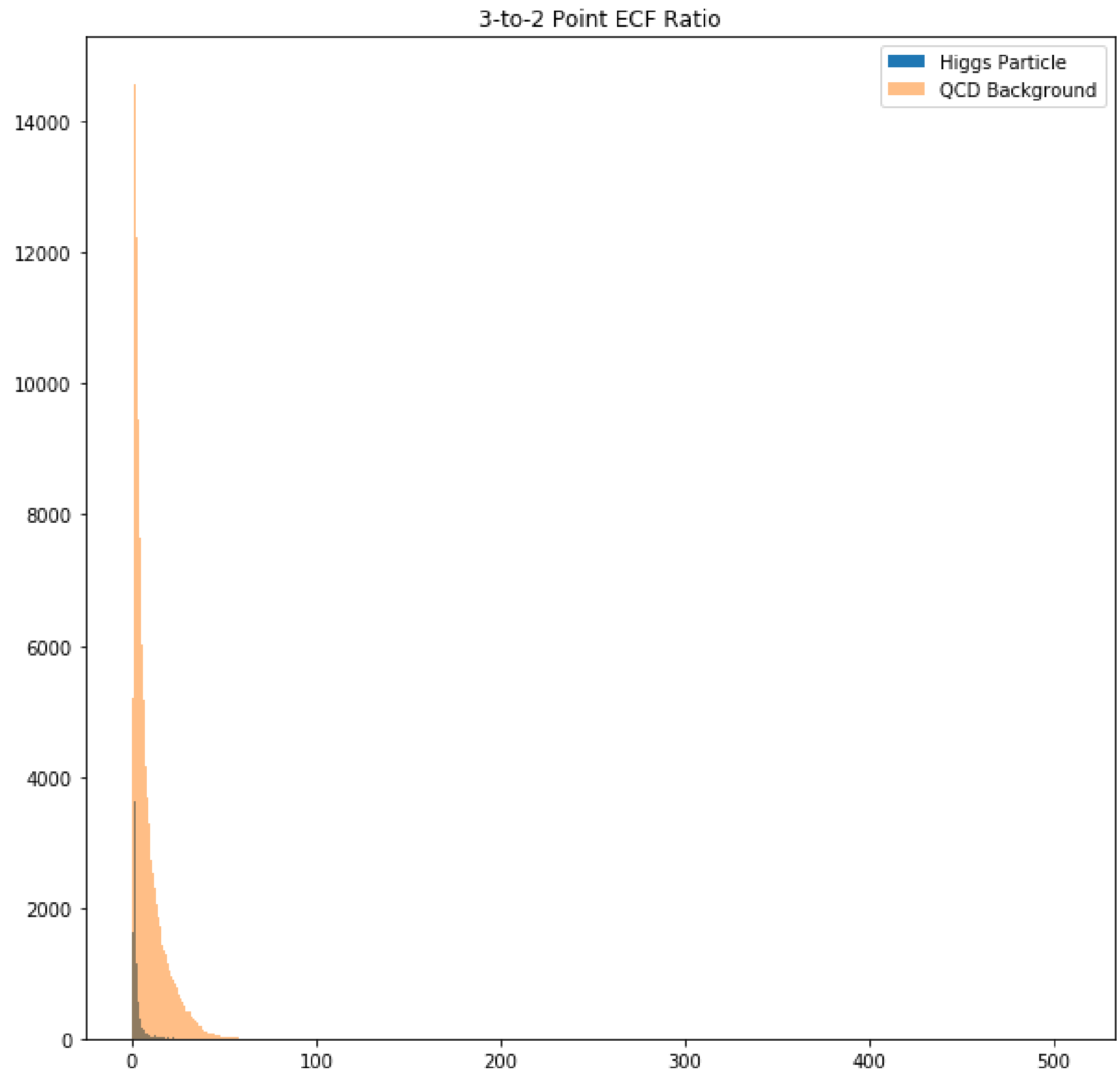
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsEe3,500,label='Higgs Particle')
ax.hist(qcdEe3,500,alpha=.5,label='QCD Background')
ax.set_title("3-Point ECF Ratio")
ax.legend(loc='upper right')
plt.show()
```



3-to-2 Point E_{CF} Ratio

The background here makes it difficult to even see relevant data. I suspect, given this is a ratio of the quark/gluon discrimination ratio and the boson identification ratio, this is showing a good representation of how full of non-interesting (for our study) decays the QCD background is. It also demonstrates just how small of a chance we have of actually identifying a boson in this manner.

In [16]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsD2,500,label='Higgs Particle')
ax.hist(qcdD2,500,alpha=.5,label='QCD Background')
ax.set_title("3-to-2 Point ECF Ratio")
ax.legend(loc='upper right')
plt.show()
```



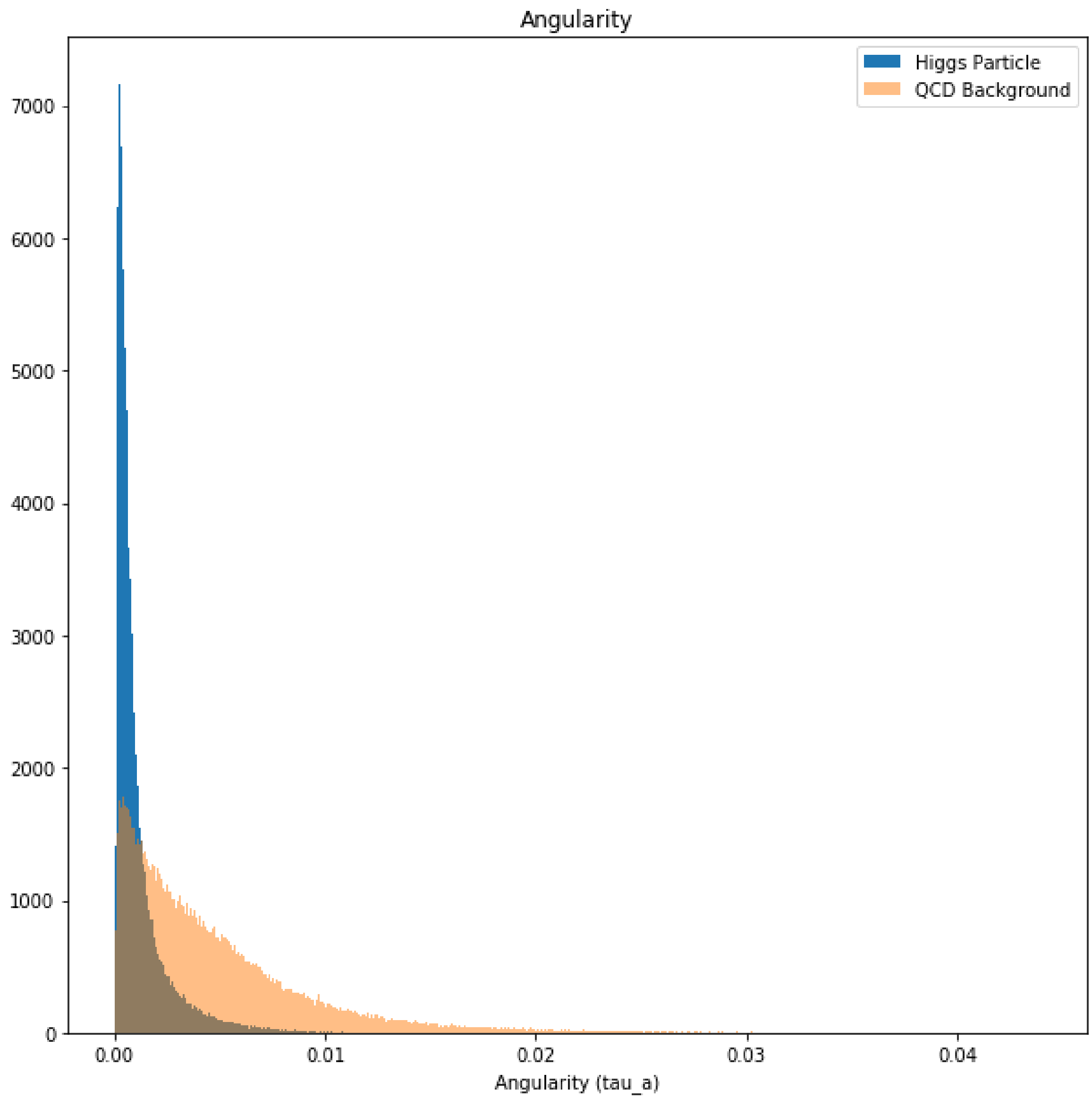
Angularity

The distribution correlates the lower limit to symmetric decay, while the upper limit corresponds to a very asymmetric decay.

As previously noted, angularity is suggested as a separator between QCD and heavy object jets.

The histogram below confirms this suggestion, as the distribution from the QCD jets is not nearly as dramatic as the the Higgs jets.

In [31]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsAng,500,label='Higgs Particle')
ax.hist(qcdAng,500,alpha=.5,label='QCD Background')
ax.set_title("Angularity")
ax.legend(loc='upper right')
ax.set_xlabel("Angularity (tau_a)")
plt.show()
```



N-Subjettiness

As previously mentioned, N-subjettiness is defined by:

$$\tau_N = \frac{1}{d_0} \sum_{k=1}^{M} (p_{T,k} \times \Delta R_{min,k})$$

where $\Delta R_{min,k}$ is the distance to the nearest subject and d_0=R $\times$ sum of all p_T of all constituents

If τ_N is large, then a description in terms of > N subjects is better.

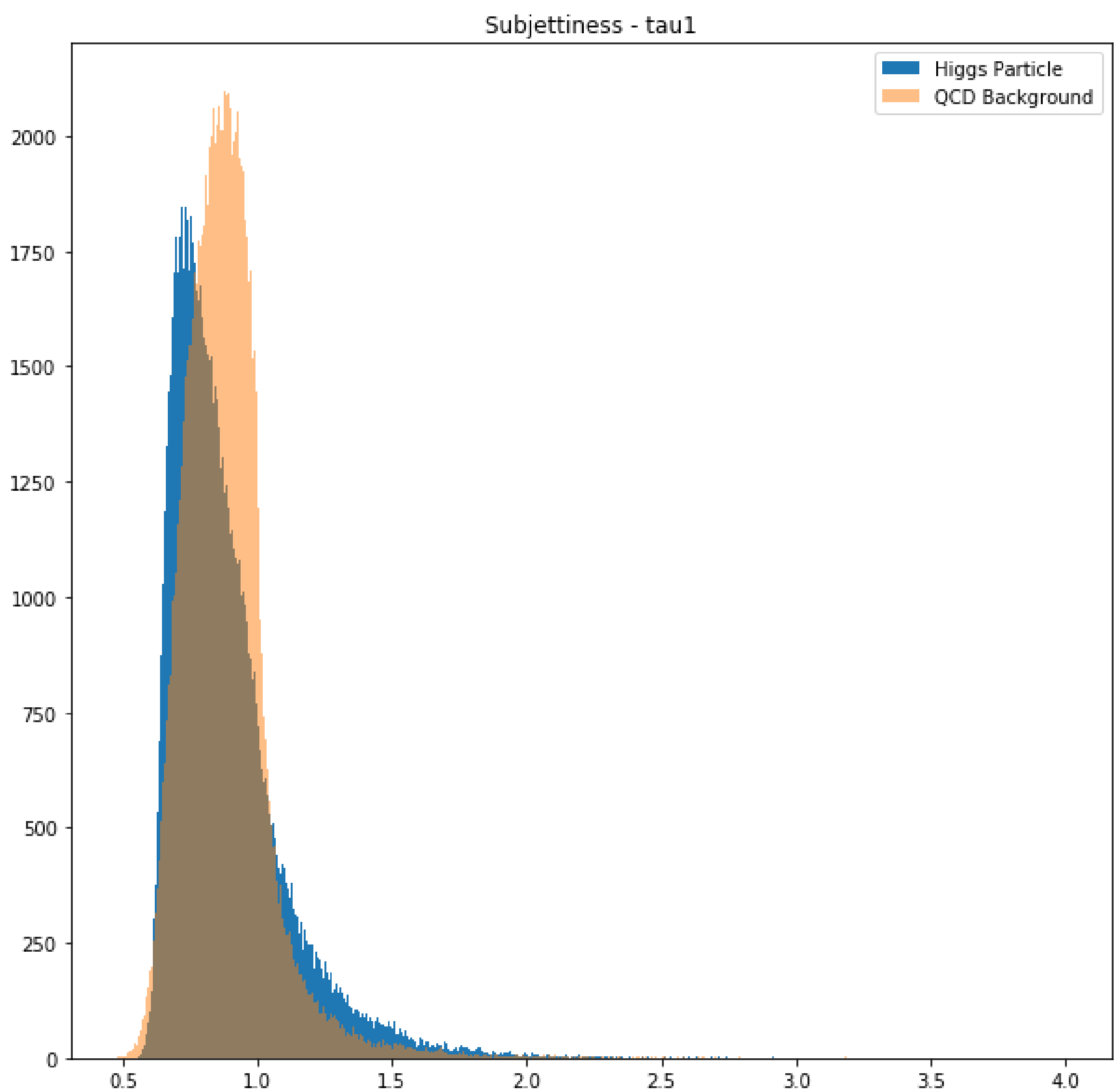
It should be noted that, to differentiate boosted objects (such as our Higgs boson), we should look at ratios. For example, Higgs will have a large $\tau 1$ and smaller $\tau 2$. For the background, it should be stated that have large $\tau 1$ are gennerally diffused, and will have larger $\tau 2$.

Also to keep in mind while looking through the below histograms, the best single discriminating variable is $\tau 2 \setminus \tau 1$

$\tau 1$

In [29]:

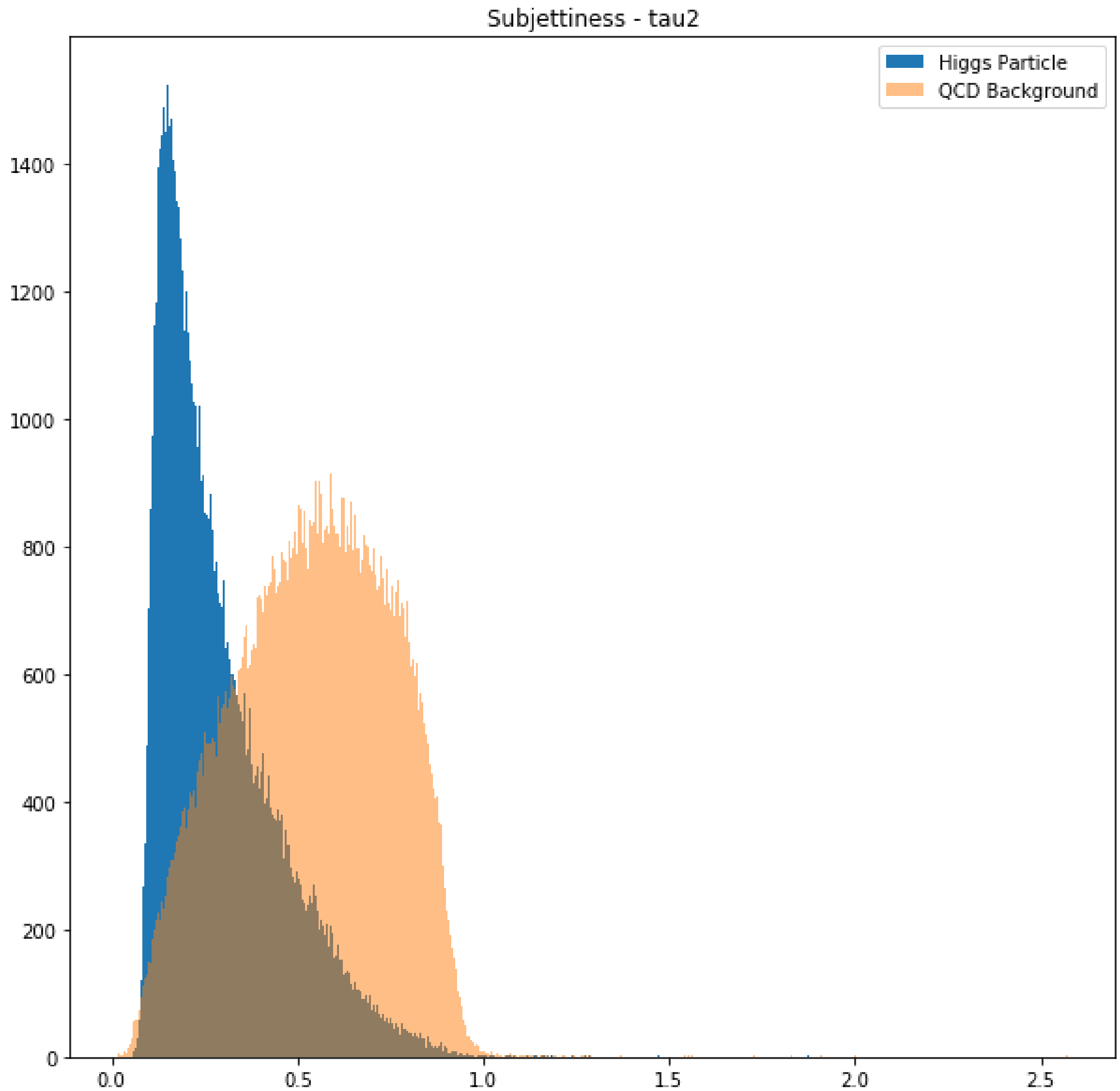
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsT1,500,label='Higgs Particle')
ax.hist(qcdT1,500,alpha=.5,label='QCD Background')
ax.set_title("Subjettiness - tau1")
ax.legend(loc='upper right')
plt.show()
```



$\tau 2$

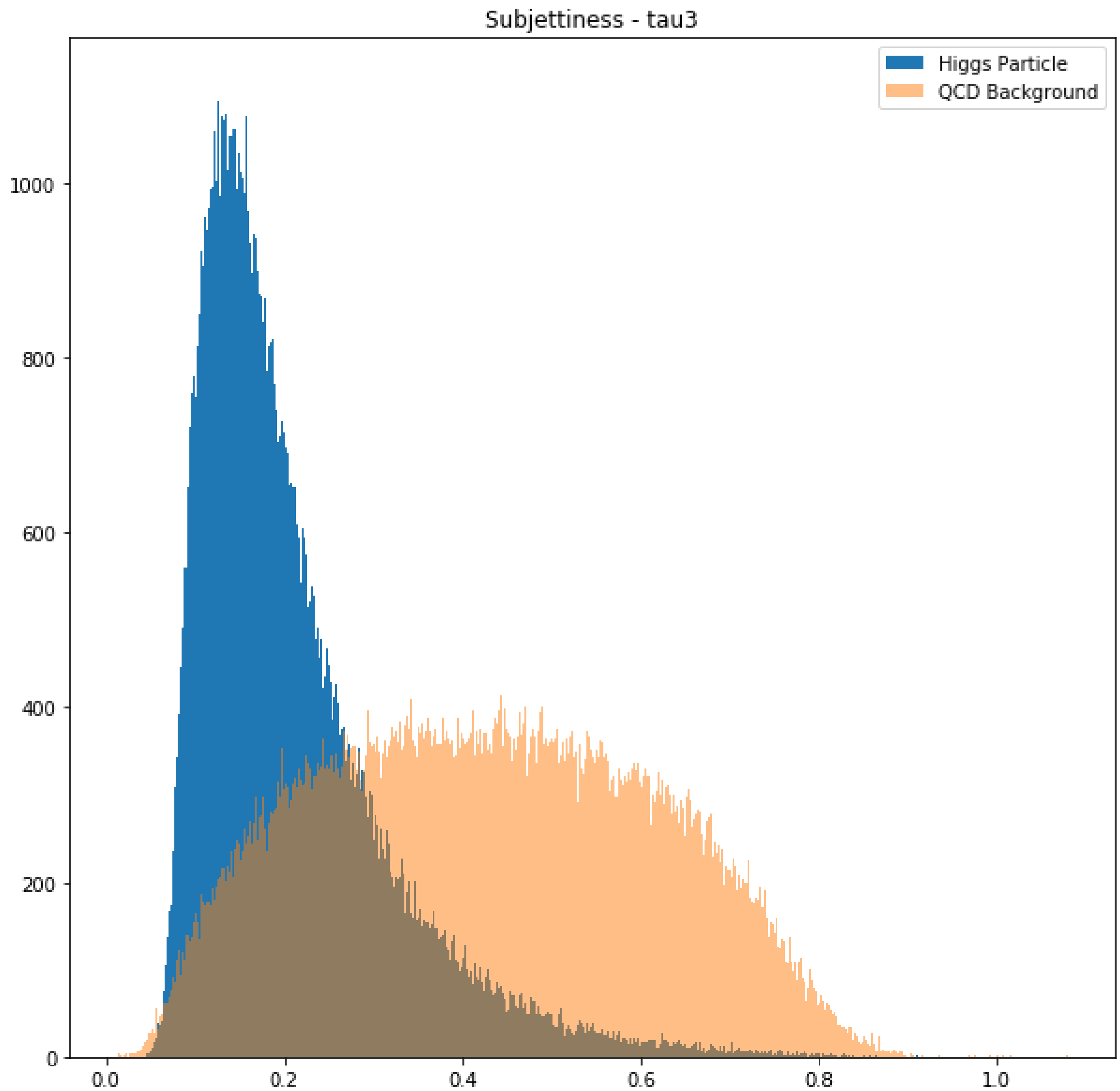
In [19]:

```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsT2,500,label='Higgs Particle')
ax.hist(qcdT2,500,alpha=.5,label='QCD Background')
ax.set_title("Subjettiness - tau2")
ax.legend(loc='upper right')
plt.show()
```



$\tau 3$

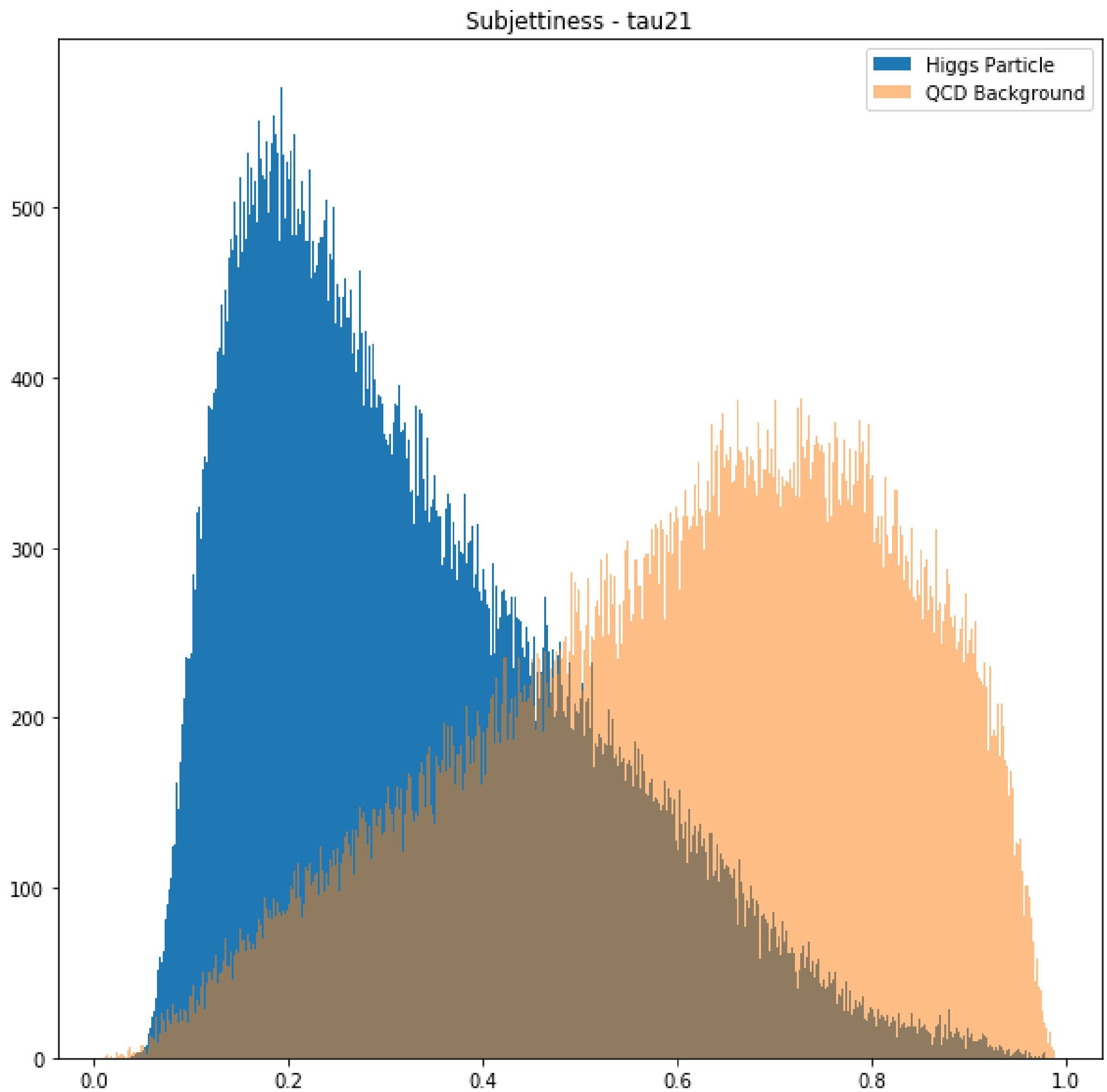
In [27]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsT3,500,label='Higgs Particle')
ax.hist(qcdT3,500,alpha=.5,label='QCD Background')
ax.set_title("Subjettiness - tau3")
ax.legend(loc='upper right')
plt.show()
```



τ21

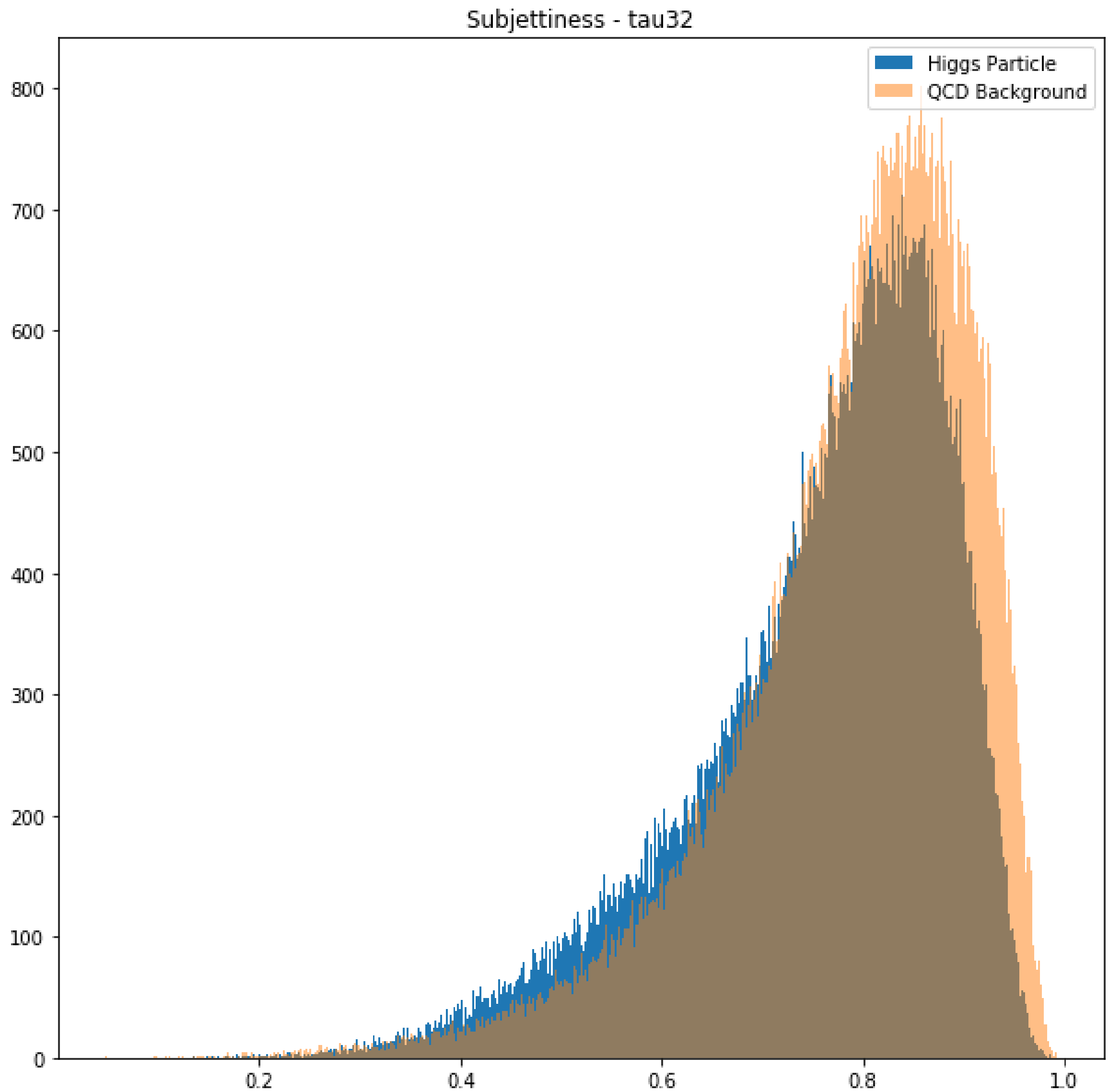
It is interesting to note the distinction between τ21 and τ32. As noted above, the best discriminator is τ21, which is clear looking at the histogram visualizations of the two.

In [21]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsT21,500,label='Higgs Particle')
ax.hist(qcdT21,500,alpha=.5,label='QCD Background')
ax.set_title("Subjettiness - tau21")
ax.legend(loc='upper right')
plt.show()
```



τ32

In [23]:
```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsT32,500,label='Higgs Particle')
ax.hist(qcdT32,500,alpha=.5,label='QCD Background')
ax.set_title("Subjettiness - tau32")
ax.legend(loc='upper right')
plt.show()
```



KtDeltaR

So far, I've been unable to really understand what this is. kT appears to be some kind of jet-finding algorithm, and KtDeltaR is another algorithm that uses kT with different jet radii to find... something.

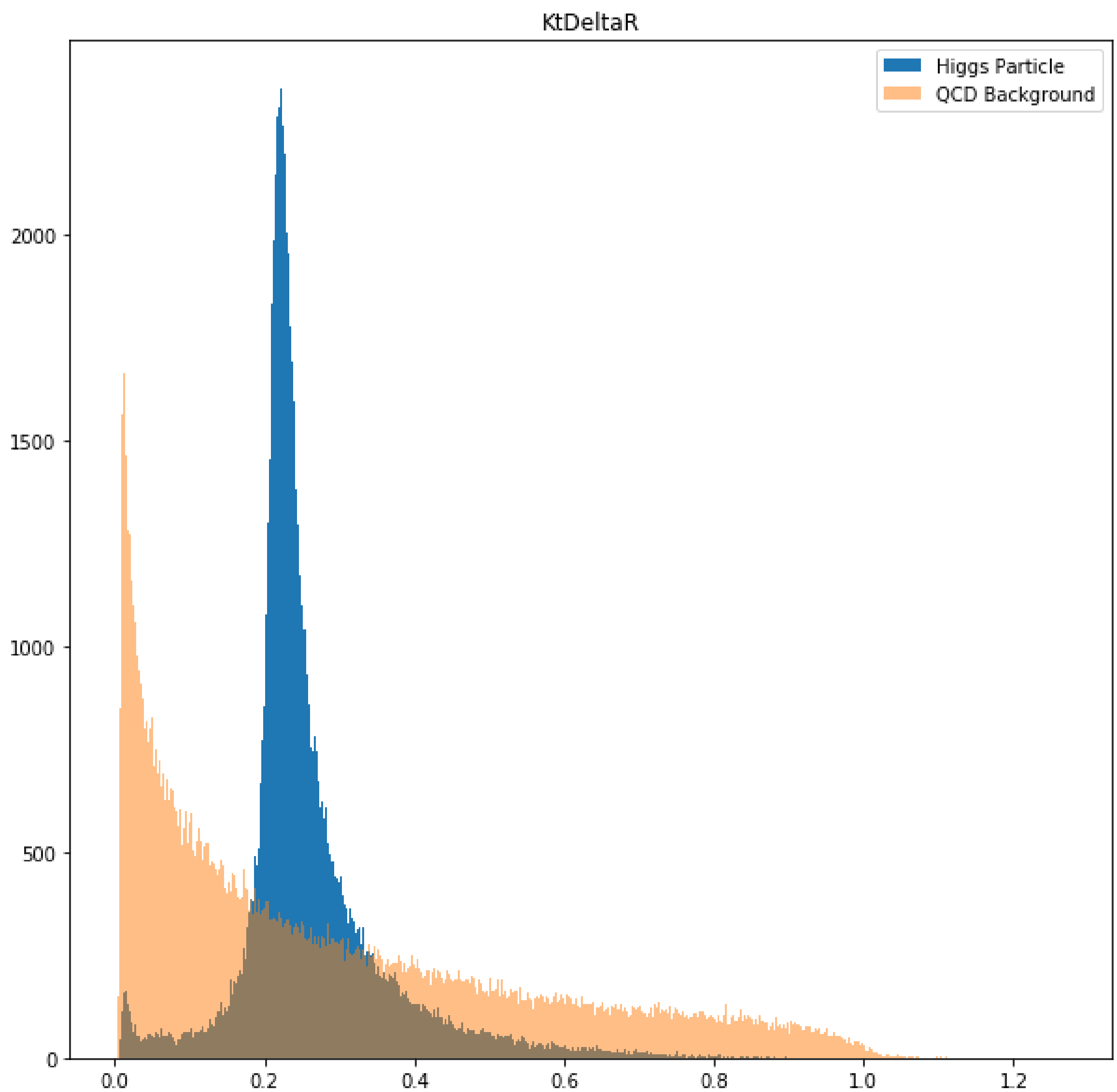
With that being said, there are a couple of interesting features in the below histogram.

There is an obvious peak of the Higgs data that exceeds the QCD background at about .25. Depending on how this is significant, if it is significant, it would be very easy to use this comparison for future calculations.

In addition, there is a small peak in the Higgs data at nearly 0.0. If this is a significant feature, we would not be easy to 'find' it because of the significant background noise.

In [33]:

```
fig, ax = plt.subplots(1, 1, figsize=(10,10))
ax.hist(higgsKtD,500,label='Higgs Particle')
ax.hist(qcdKtD,500,alpha=.5,label='QCD Background')
ax.set_title("KtDeltaR")
ax.legend(loc='upper right')
plt.show()
```



Conclusion

Some of the data above is not fully understood, and so the significance of the visualizations is lost (such at KtDeltaR).

However, it is clear that certain features of the dataset will be important in optimizing the QCD background data.

If I were to continue this process, without further knowledge about the data or the significance of what I've observed with the above histograms, I would find the following features to be more important than the rest:

- **Invariant Mass**
  - Chosen due to the clear peak at 125 GeV, which is the known mass of the Higgs boson
- **Angularity**
  - Chosen because of the clear peak the Higgs data has over the QCD background data. In addition, there are useful data in regards to decays given an upper and lower bound in the Higgs distribution.
- τ**21**
  - Is a suggested excellent jet discriminator between background and boosted object jets. This suggestion is validated based on the comparison between $\tau 21$ and $\tau 32$

Resources Used in Data Exploration:

These were sources not provided in the lab instructions themselves

Energy Correlation Functions for Jet Substructure (https://arxiv.org/abs/1305.0007)
Computing N-subjettiness for boosted jets (https://arxiv.org/abs/1809.04602)
TASI Lectures on Jet Substructure (https://arxiv.org/pdf/1302.0260.pdf)
Weizmann Institute of Science
(https://indico.cern.ch/event/170556/contributions/1430595/attachments/211988/297161/pic2012_du
ALICE Publications (http://alice-publications.web.cern.ch/node/3701)

In [ ]: