# TABLE OF CONTENTS

[**TABLE OF CONTENTS**](#_ft094pd0yk7r) **1**

[**Datasets**](#_jne7rc4wc62l) **1**

[Urban Forests](#_5g1a8vrrjgt9) 1

[eBirds](#_9tj3vcqkb8du) 2

**Hexagon** [**Grid Approximation**](#_bqxk65j4gjfn) **5**

[**Hexagon Grid Approximation**](#_swv5w0xn2dii) **6**

[**Graphical Plots**](#_5cztdh6dw2u6) **7**

# Datasets

Jackie Points:

* Diversity:
  + Evenness, are they single individuals?
    - Shannon-Wiener Index
  + Number of species that make up the cumulative 50% of individuals
    - Min-num of species to make 50%
  + Maybe look into Entropy? Pretty much same as Shannon-Index
* Birds are very particular to areas
  + Only hang out in certain areas that are small
* Find the Goldilocks size for grids
  + Get many different sizes
  + Pick some measures
  + Find the smallest sizes that doesn’t change the shape/pattern
    - Or which side is best when it changes
* Hexagons
  + Can be used when using distance between centroids
  + All neighbors are equidistant to the center
* Investigate Bird-Watching around Covid Lockdowns

Ventura Points

* Color bar => Color matrix for both datasets on California graph
* Move to make a graph of Tree and Bird Data together
* Virtual Table??
* His Project
  + U-Net
  + VGG Encoder / Decoder ahead

## Urban Forests

* Uploaded the Native Tree lists into BigQuery
* Renamed `final` dataset to `final-without-nativity`
* Added Indicator column `nativity` to `final` dataset
* *Lots of species don’t include epithet but our nativity dataset does (Concern)*
* native\_df contains the unique native tree species to California
* tree\_species contains all the unique tree species in the urban tree dataset
* 38 Native Species seem to not show up in our tree dataset (Investigation)
  + Over 100 **Abies** with no Epithet. None with epithet are labeled “**amabilis**”
    - Amabilis is barely native to California (Extreme North West)
  + **Arctostaphylos** **Manzanita** is mis-spelled in the Native List ✓
  + **Alnus Incana** has a subspecies that is sparsely native to North California but mainly Canada/Alaska
  + **Alnus Viridis** is an [illegitimate](https://en.wikipedia.org/wiki/Alnus_alnobetula) name for another species, that also doesn’t exist in the tree dataset
  + **Ceanothus Thyrsiflorus** is mis-spelled in the Native List ✓
  + **Celtis Reticulata** has a different name in the Tree Dataset (Celtis laevigata var. reticulata) ✓
  + **Cercocarpus Ledifolius** is mis-spelled in the Native List, but also not in Tree Data
  + **Cornus Nuttallii** is mis-spelled in the Native List ✓
  + The Cupressus genus was recently reclassified into **Hesperocyparis**(So they don’t match)
    - Only **Sargentii** exists in the Tree Dataset ✓
    - All other species are incredibly endangered and local to small areas
  + **Juniperus Grandis** is considered a subspecies of Occidentalis which is in Tree Data
  + **Juniperus Osteosperma** are barely native to California and only found in high elevation
  + **Lyonothamnus Floribundus** have a much longer name in Tree Data (Lyonothamnus floribundus subsp. aspleniifolius) ✓
  + **Parkinsonia Microphylla** is barely found in South-East California
  + **Picea Breweriana** is incredibly rare
  + **Picea sitchensis** is mis-spelled in the Native List ✓
  + **Pinus Jeffreyi** is mis-spelled in the Native List ✓
  + **Pinus Washoensis** is a subspecies of Pinus Ponderosa
    - All other missing **Pinus**’s are found in high elevation
  + **Populus Angustifolia** is barely found in California
  + **Prunus Ilicifolia subsp. lyonii**: The L isn’t capitalized in Tree Dataset ✓
  + **Quercus Wislizen**i is mis-spelled in the Native List ✓
  + **Quercus Parvula** isn’t in Tree Dataset but is taxonomically very similar to **Wislizeni**
  + **Salix Scouleriana** is only found in the Sierra Nevada range in California
  + **Sequoiadendron Giganteum** is mis-spelled in the Native List ✓
  + **Staphylea Bolanderi** is only found in the Sierra Nevada range in California
  + **Torreya Californica** is mis-spelled in Native List ✓
  + **Tsuga Mertensiana** is found in snow/high elevation. Tree Data has lots of Tsugas that could be this
* Appended a grid id column to the final dataset that indicates which hexagon the tree belongs to
  + Only 530 hexagons have observed tree data

## eBirds

**LIST OF DROPPED COLUMNS**

Dropped columns with 64 million + nulls

* *typeStatus*
* *depth*
* *depthAccuracy*
* *coordinatePrecision*
* *establishmentMeans*
* *recordNumber*
* *verbatimScientificNameAuthorship*
* *elevationAccuracy*
* *elevation*
* *infraspecificEpithet*
* *mediaType*
* *coordinateUncertaintyInMeters*
* *rightsHolder*
* *dateIdentified*
* *identifiedBy*
* *issue*

Dropped columns not relevant for our project

* *datasetKey*, because we are not concerned with dataset origin
* *occurrenceID*, because gbifID is already a unique identifier
* *taxonKey*
* *speciesKey*
* *institutionCode*
* *collectionCode,* all ebird
* *catalogNumber*
* *license*
* *recordedBy*
* *lastInterpreted,* this is a date, but we have eventDate, which is date of observation
* *scientificName,* is just verbatimScientificName with creator and date
* *species*, is the same as verbatimScientificName with more nulls
* *countryCode*, all observations in US
* *stateProvince*, all observations in California

We filtered to animalia-chordata-aves only, therefore we dropped

* *kingdom*
* *phylum*
* *class*

filtered to human observation or observation only for basis of record

* *basisOfRecord*

all occurrenceStatus is labeled as present

* *occurrenceStatus*

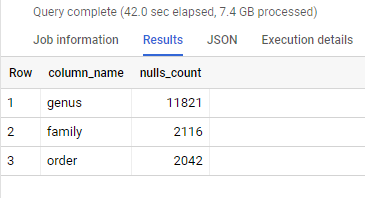
have lat long coordinates

* *locality*

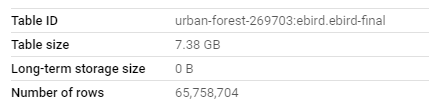
Mapped individualCount to be 1 if null, original value otherwise

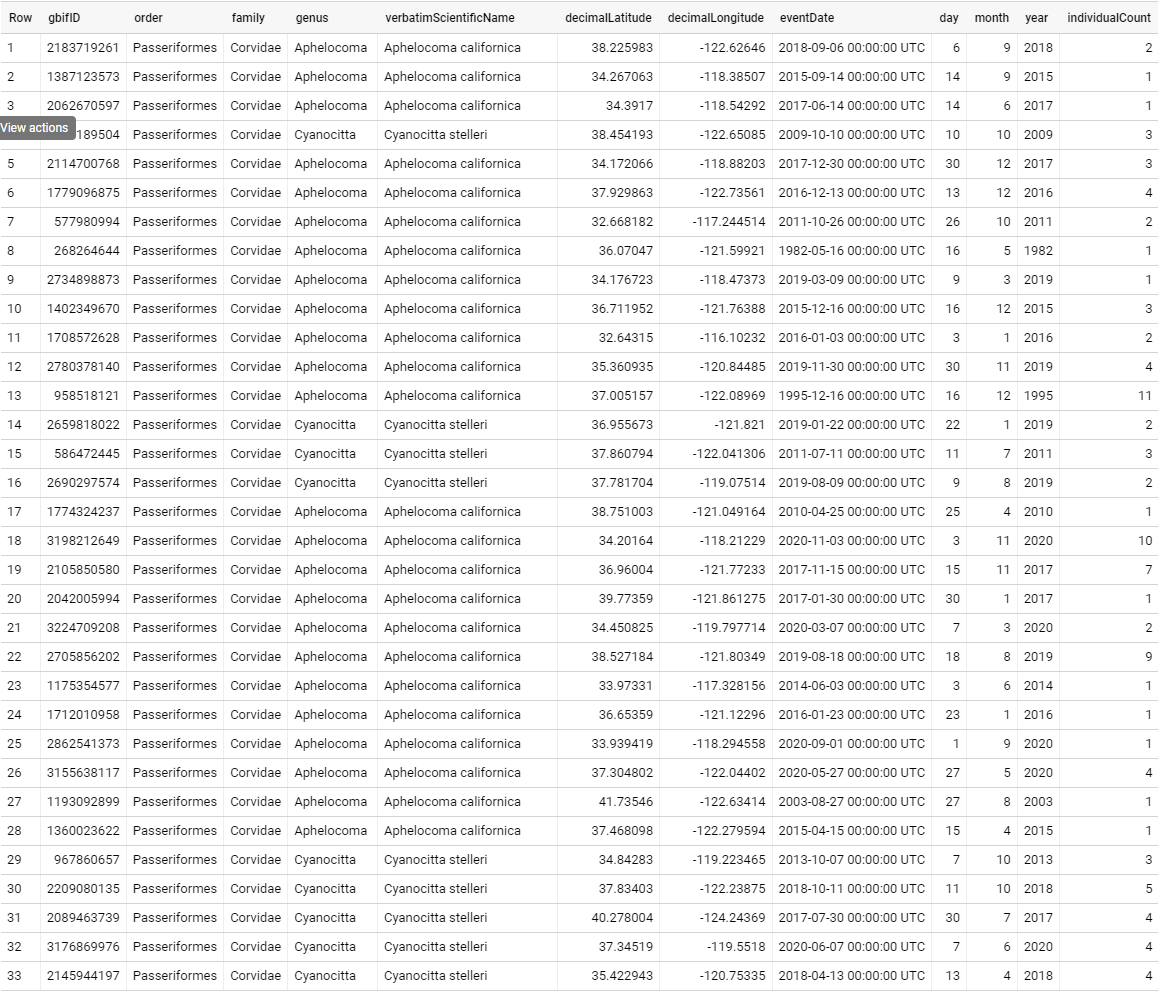
Filtered years to be between 1980 and 2021

Remaining null count per column below

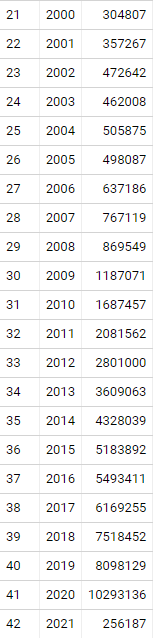
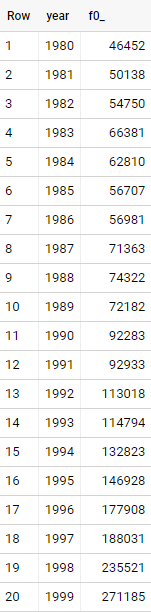


* Final Table Preview





number of observations per year



# 

# 

# 

# 

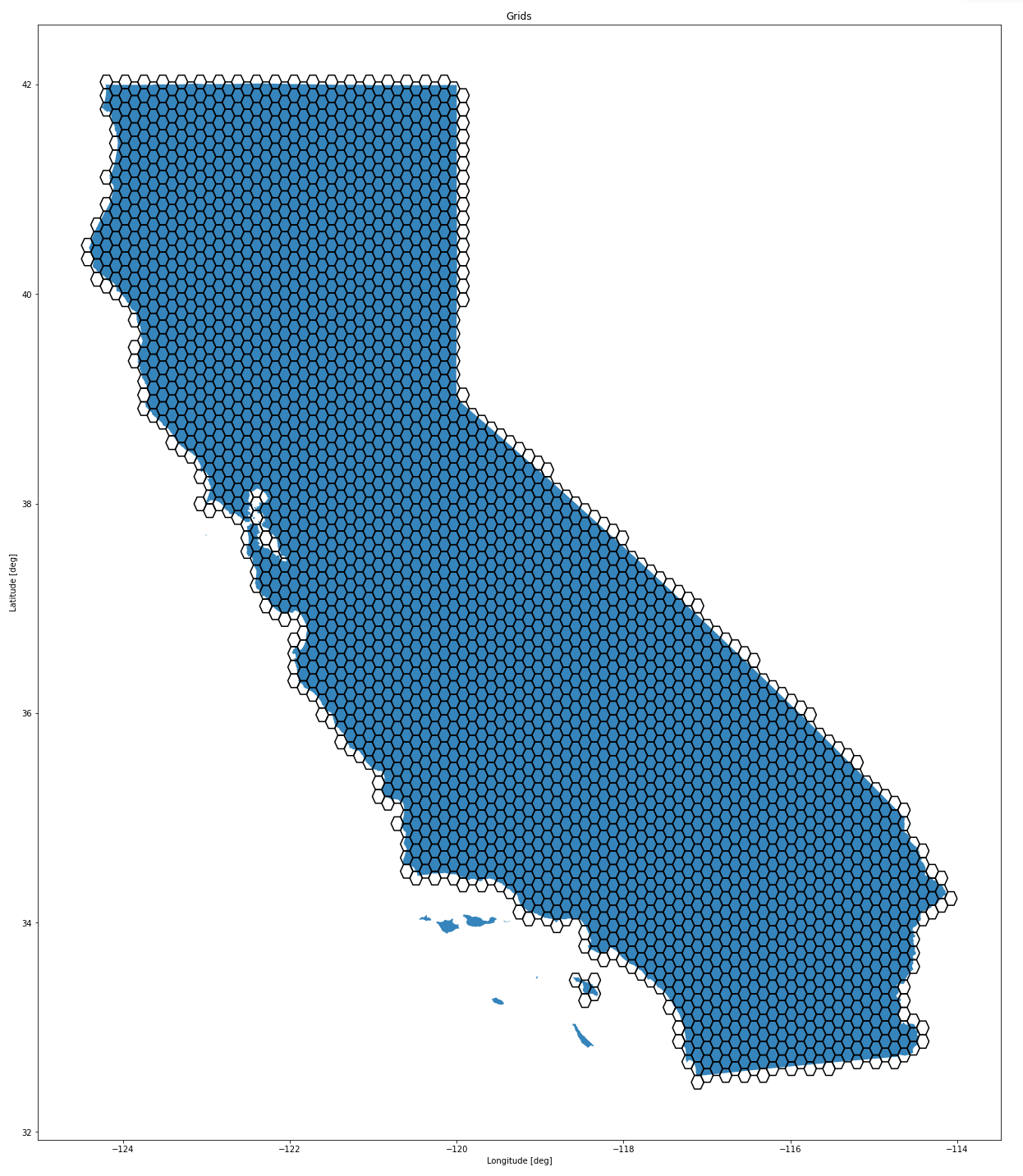
# 

# 

# 

# 

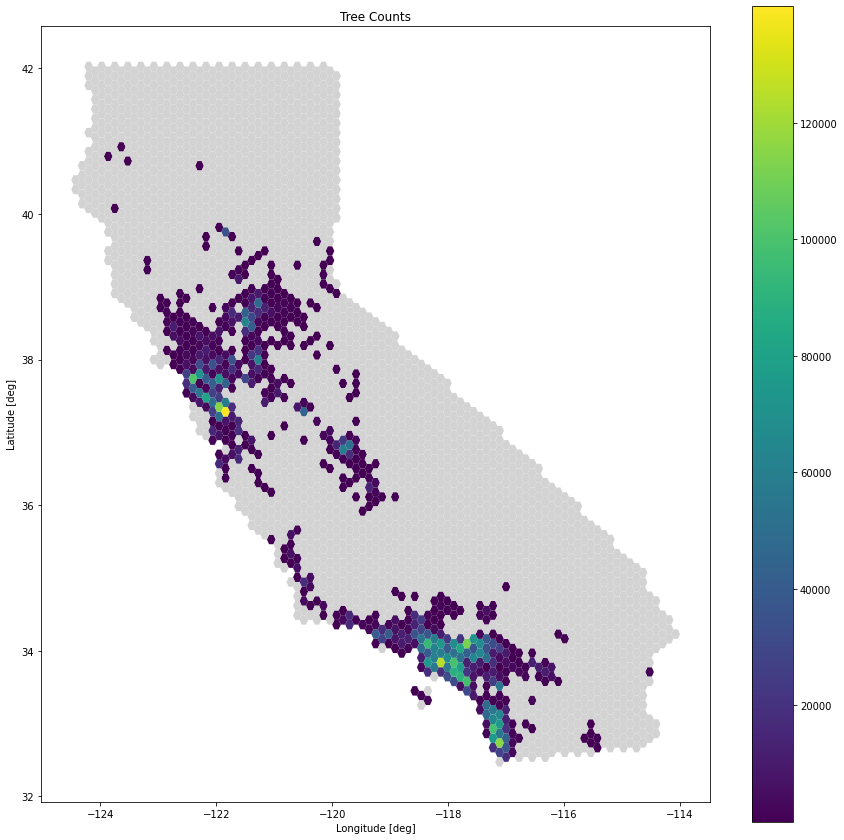
**Hexagon Grid**

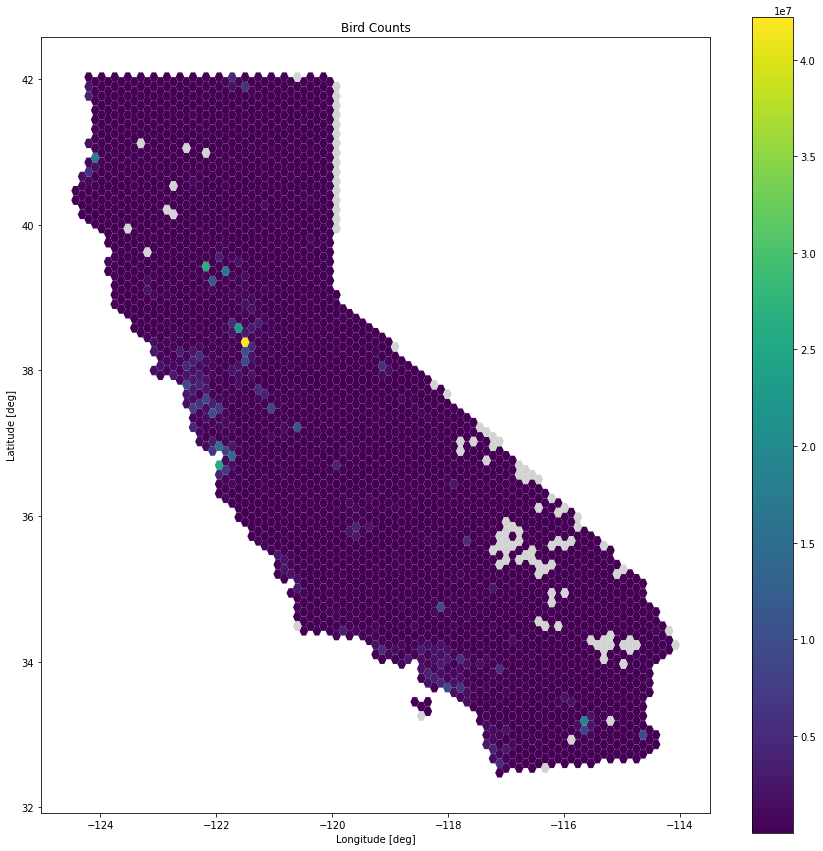
3029 Hexagons, approximately 53km^2 each

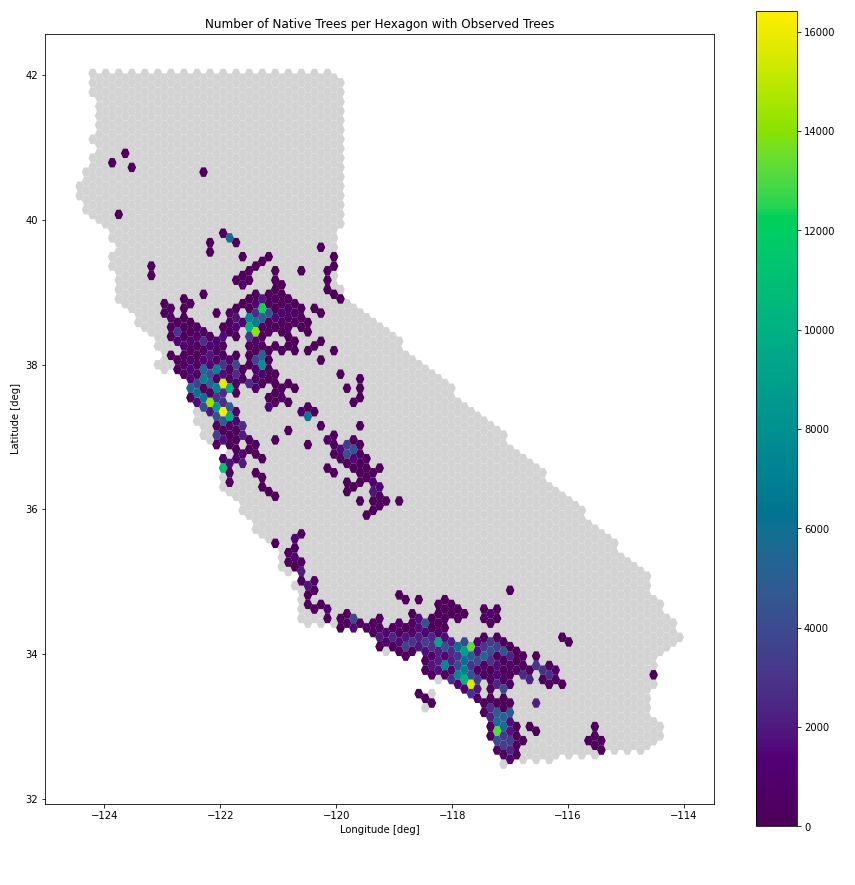
Appended a grid id column to the final dataset that indicates which hexagon the tree belongs to

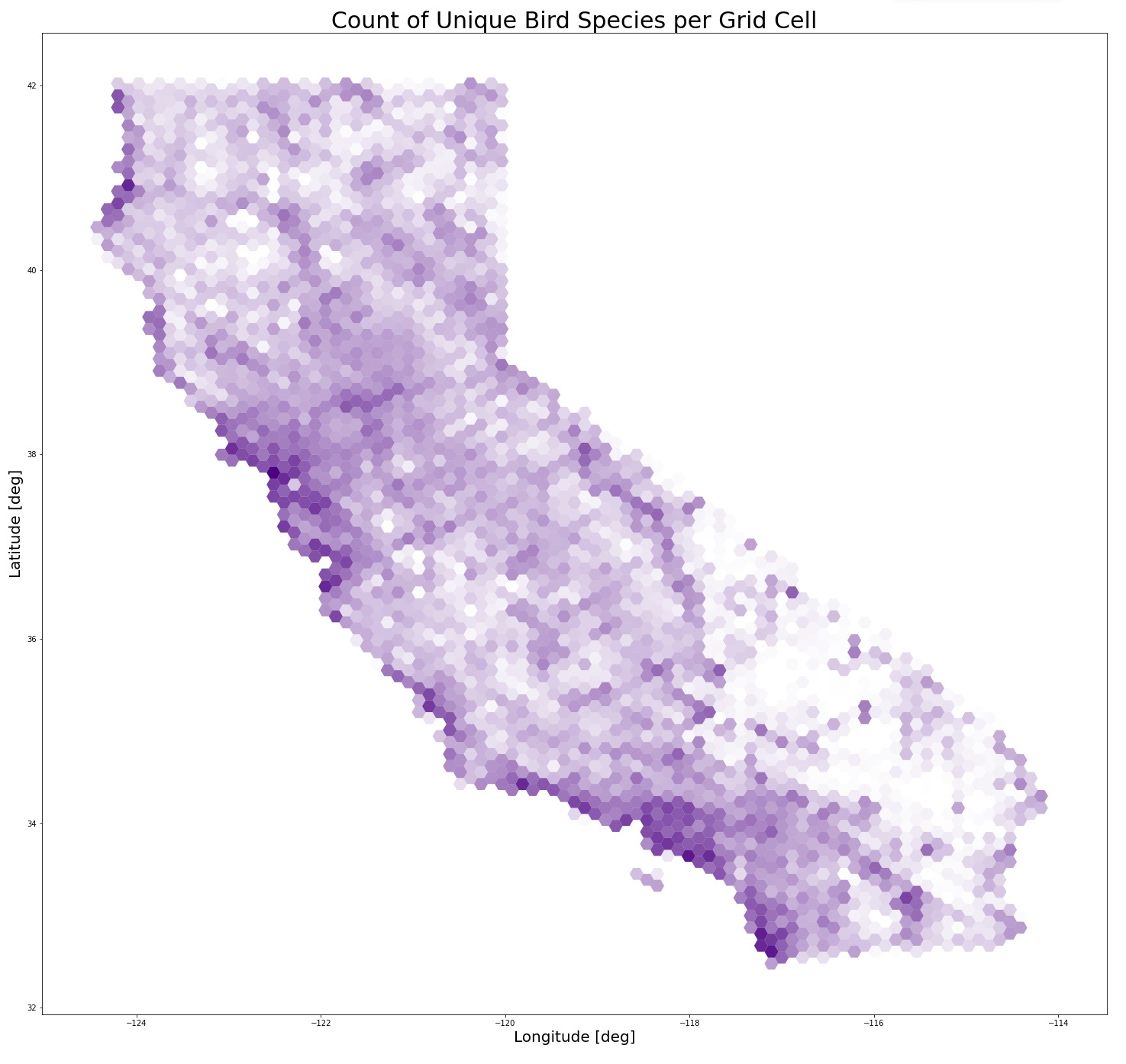
# Graphical Plots

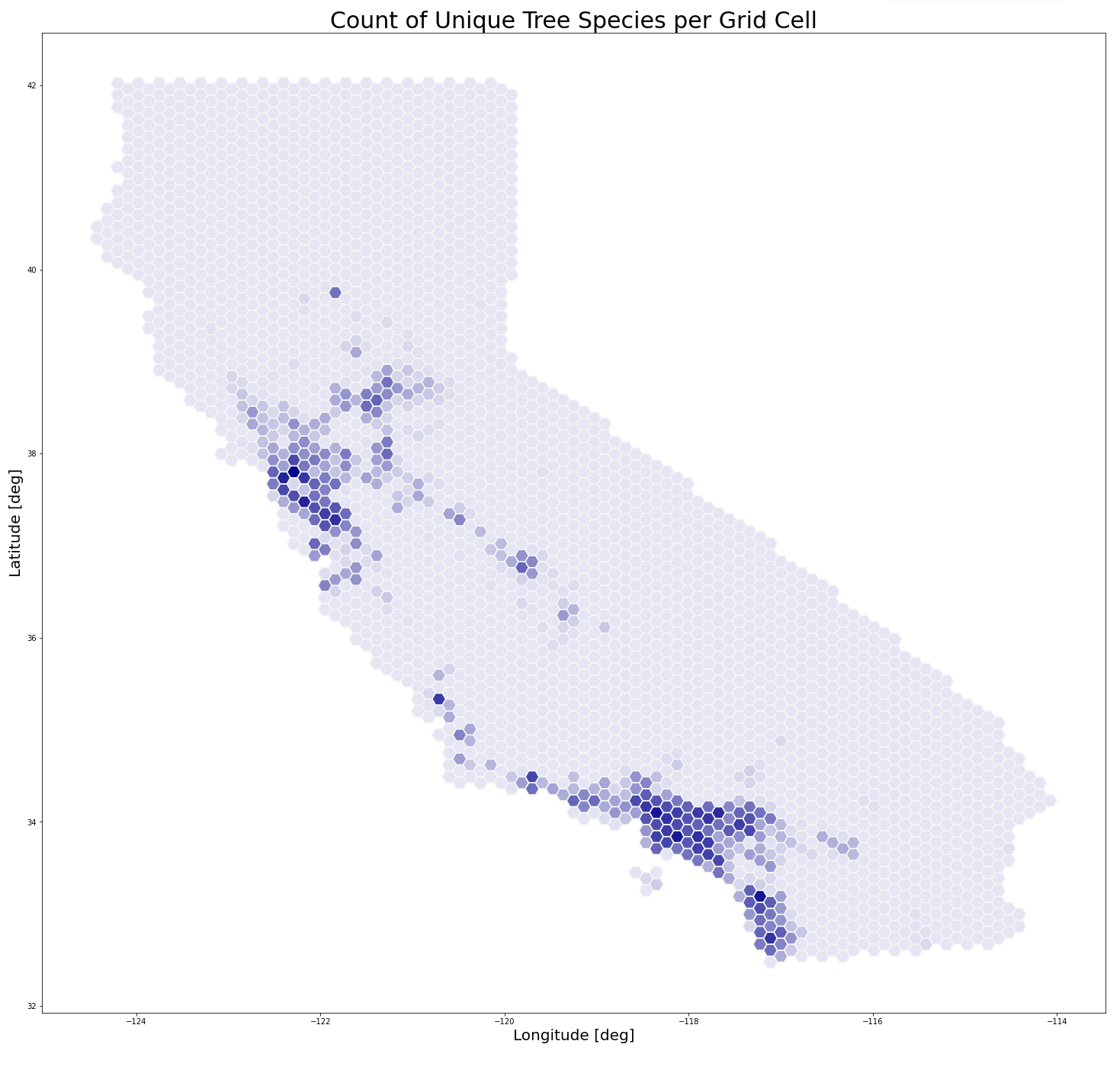
Tree Counts by Hexagon on Entire Dataset



Bird Counts by Hexagon (notice scale)







Shannon-Wiener Index for Birds

