

# Final\_Project

March 23, 2017

Hyein (Haily) Kil  
804268393  
EEB C177 Final Project

## 1 Introduction

Species richness is decreasing at a global scale due to man-made and natural factors (Sax, et al. 2002). To compensate for the changes in the environment, new species are introduced to the land either by speciation or migration, which further develops the phylogenetic tree. The phylogenetic tree can continue to be updated by extensive research on fossil records that can often discover relationships between different species that seem unrelated (Sorenson, et al. 1999).

It is discovered that *Anas acuta*, a bird species native to North America, carries an Asian-origin influenza strain (Koehler, et al. 2008), illustrating a relationship between American bird and Asian bird.

In this project, I will observe the occurrences of fossil records of *Anas acuta* in respective to the fossil records of *Anas platyrhynchos* (Asian-origin bird) to see whether I can draw some conclusion about the relationship between the two species of birds.

I will be using the csv file that contains the list of North American native bird species from the American Birding Association ([listing.aba.org/aba-checklist](http://listing.aba.org/aba-checklist)). The list of the fossil records is taken from the Paleobiology Database (<https://paleobiodb.org>).

Here is the link to my github address: <https://github.com/hailykil/eeb-177-final-project>

```
In [1]: # Picture of Anas acuta - North American origin
        from IPython.display import Image
        Image(filename = "acuta.png")
```

Out [1]:



Here is the picture of the bird native to North America. This is the species of bird that is discovered to carry the Asian-origin influenza strain.

```
In [3]: # Picture of Anas platyrhynchos - Asian origin
        Image(filename = "chinese_duck.png")
```

Out [3]:



And this is the picture of the Asian-origin bird. This is the species of bird that I will be comparing to *Anas acuta*.

## 2 Python

### 2.1 Code Part 1

Step 1. First, I created a dictionary that listed the unique Anatidae species to the time interval where the fossil records were found.

```
In [16]: # function 1: create dictionary
# Unique species : time interval where the fossil records were found

# makes a new function called makebirdDict
def makebirdDict(filename):

    birdDict = {} # creates an empty dictionary
    fobj = open(filename, "r",
                  encoding = "ISO-8859-15") # opens the file
    # read the file starting from line 1
    all_records = fobj.readlines()[1:]
    for line in all_records:
        # removing unnecessary empty lines
        line = line.strip().strip('\n')
        # split the lines due to delimiters
        record_elements = line.split(",")
        # read the first column of the file
        occurrence = record_elements[0]
        # read the second column of the file
        species = record_elements[1]
        # matches occurrence and species
        birdDict[occurrence] = species

    return birdDict # return dictionary
```

Step 2. Then, I used the previously formed dictionary and took out lines that only contained the species that I am interested. In this case, it would be *Anas acuta* and *Anas platyrhynchos*.

```
In [18]: # Recall previously formed dictionary...
bird = makebirdDict("bird_data.csv")
occurrence = list(bird.keys())
bird
# Only take out Anas acuta
{k:v for k,v in bird.items() if 'acuta' in v}

Out[18]: {'"1083779"': '"Anas acuta"',
          '"1196606"': '"Anas acuta"',
          '"1197460"': '"Anas acuta"',
```

```

"1197716": "Anas acuta",
"1201083": "Anas acuta",
"1313401": "Anas acuta",
"213766": "Anas acuta",
"219352": "Anas acuta",
"219392": "Anas acuta",
"219498": "Anas acuta",
"219518": "Anas acuta",
"447659": "Anas acuta",
"594523": "Anas acuta",
"623927": "Anas acuta",
"665511": "Anas acuta",
"690674": "Anas acuta",
"841379": "Anas acuta"}

```

This is the dictionary that only took out lines that contained “Anas acuta”

```

In [17]: # Recall the previously formed dictionary...
bird = makebirdDict("bird_data.csv")
occurrence = list(bird.keys())
bird
# Only take out Anas platyrhynchos
{k:v for k,v in bird.items() if 'platyrhynchos' in v}

```

```

Out[17]: {"1020839": "Anas platyrhynchos",
"1024172": "Anas platyrhynchos",
"1024173": "Anas platyrhynchos",
"1186504": "Anas platyrhynchos",
"1195901": "Anas platyrhynchos",
"1196100": "Anas platyrhynchos",
"1196603": "Anas platyrhynchos",
"1197044": "Anas platyrhynchos",
"1197459": "Anas platyrhynchos",
"1197712": "Anas platyrhynchos",
"1197895": "Anas platyrhynchos",
"1201082": "Anas platyrhynchos",
"1313400": "Anas platyrhynchos",
"1314658": "Anas platyrhynchos",
"1318491": "Anas platyrhynchos",
"213777": "Anas platyrhynchos",
"213817": "Anas platyrhynchos",
"219357": "Anas platyrhynchos",
"219457": "Anas platyrhynchos",
"219504": "Anas platyrhynchos",
"219524": "Anas platyrhynchos",
"370509": "Anas platyrhynchos",
"447661": "Anas platyrhynchos",

```

```

'"594522"' : '"Anas platyrhynchos"',
'"665510"' : '"Anas platyrhynchos"',
'"709981"' : '"Anas platyrhynchos"',
'"710085"' : '"Anas platyrhynchos"',
'"740024"' : '"Anas platyrhynchos"',
'"740518"' : '"Anas platyrhynchos"',
'"781719"' : '"Anas platyrhynchos"',
'"789196"' : '"Anas platyrhynchos"',
'"840107"' : '"Anas platyrhynchos"',
'"841378"' : '"Anas platyrhynchos"',
'"915167"' : '"Anas platyrhynchos"'
}

```

This is the dictionary that only took out lines that contained “Anas platyrhynchos”

As you can see, from many fossil records catagorized under family “Anatidae,” there are numerous fossil records of Anas acuta and Anas platyryhynchos. However, simply looking at the occurrences of the fossil records of the bird species does not provide us with any information about the relationship between the two. Therefore, looking at the time intervals where the fossil records were found along with the occurrences of the fossil records is crucial in drawing conclusions between the two species.

## 2.2 Code Part 2

Step 3. I created another dictionary that matched occurrences of the fossil records to the time interval where they were found.

```

In [8]: # function 2: dictionary 2
        # Using data for Anas acuta only
        # Occurrences : Time period the fossil records were found
        # This will be used to plot a bar graph

        # makes a new function called makebirdDict
def makebirdDict2(filename):
    birdDict2 = {} # creates an empty dictionary
    # opens the file
    fobj = open(filename, "r", encoding = "ISO-8859-15")
    # read the file starting from line 1
    all_records = fobj.readlines()[1:]
    for line in all_records:
        # removing unnecessary empty lines
        line = line.strip().strip('\n')
        # split the lines due to delimiters
        record_elements = line.split(",")

```

```

# read the first column of the file
occurrence = record_elements[0]
# read the fourth column of the file
interval = record_elements[3]
# matches the species to the interval
birdDict2[occurrence] = interval

# return dictionary
return birdDict2

```

```

In [19]: acuta = makebirdDict2("acuta.csv")
# Recall dictionary 2 and print dictionary
acuta

```

```

Out[19]: {'1083779': 'Rancholabrean',
'1196606': 'Late Pleistocene',
'1197460': 'Late Pleistocene',
'1197716': 'Late Pleistocene',
'1201083': 'Late Pleistocene',
'1313401': 'Late Pleistocene',
'213766': 'Pleistocene',
'219352': 'Pleistocene',
'219392': 'Pleistocene',
'219498': 'Pleistocene',
'219518': 'Pleistocene',
'447659': 'Middle Pleistocene',
'594523': 'Zanclean',
'623927': 'Rancholabrean',
'665511': 'Late Pleistocene',
'690674': 'Late Pleistocene',
'841379': 'Late Pleistocene'}

```

This is the dictionary corresponding to the species *Anas acuta*.

```

In [10]: # function 2: dictionary 3
# Using data for Anas platyrhynchos only
# Occurrences : Time period the fossil records were found
# This will be used to plot a bar graph

# makes a new function called makebirdDict
def makebirdDict3(filename):
    # creates an empty dictionary
    birdDict3 = {}
    # opens the file
    fobj = open(filename, "r", encoding = "ISO-8859-15")
    # read the file starting from line 1

```

```

all_records = fobj.readlines()[1:]

for line in all_records:
    # removing unnecessary empty lines
    line = line.strip().strip('\n')
    # split the lines due to delimiters
    record_elements = line.split(",")
    # read the first column of the file
    occurrence = record_elements[0]
    # read the fourth column of the file
    interval = record_elements[3]
    # matches the species to the interval
    birdDict3[occurrence] = interval

    # Return dictionary
    return birdDict3

```

```

In [11]: plat = makebirdDict3("plat.csv")
         plat

```

```

Out[11]: {'"1020839"' : '"Pleistocene"',
          '"1024172"' : '"Pleistocene"',
          '"1024173"' : '"Pleistocene"',
          '"1186504"' : '"Holocene"',
          '"1195901"' : '"Blancan"',
          '"1196100"' : '"Late Pleistocene"',
          '"1196603"' : '"Late Pleistocene"',
          '"1197044"' : '"Irvingtonian"',
          '"1197459"' : '"Late Pleistocene"',
          '"1197712"' : '"Late Pleistocene"',
          '"1197895"' : '"Late Pleistocene"',
          '"1201082"' : '"Late Pleistocene"',
          '"1313400"' : '"Late Pleistocene"',
          '"1314658"' : '"Late Pliocene"',
          '"1318491"' : '"Holocene"',
          '"213817"' : '"Pleistocene"',
          '"219357"' : '"Pleistocene"',
          '"219457"' : '"Pleistocene"',
          '"219504"' : '"Pleistocene"',
          '"219524"' : '"Pleistocene"',
          '"370509"' : '"Early Pleistocene"',
          '"447661"' : '"Middle Pleistocene"',
          '"594522"' : '"Zanclean"',
          '"665510"' : '"Late Pleistocene"',
          '"709981"' : '"Late Pleistocene"',
          '"710085"' : '"Late Pleistocene"',
          '"740024"' : '"Late Pleistocene"',
          '"740518"' : '"Pleistocene"',

```

```
"781719": "Late Pleistocene",  
"789196": "Late Pleistocene",  
"840107": "Late Pleistocene",  
"841378": "Late Pleistocene",  
"915167": "Late Pleistocene"}
```

This is the dictionary that corresponds to *Anas platyrhynchos*.



### 3 Graph using Python Codes

```
In [13]: # Graph: occurrences vs. time period

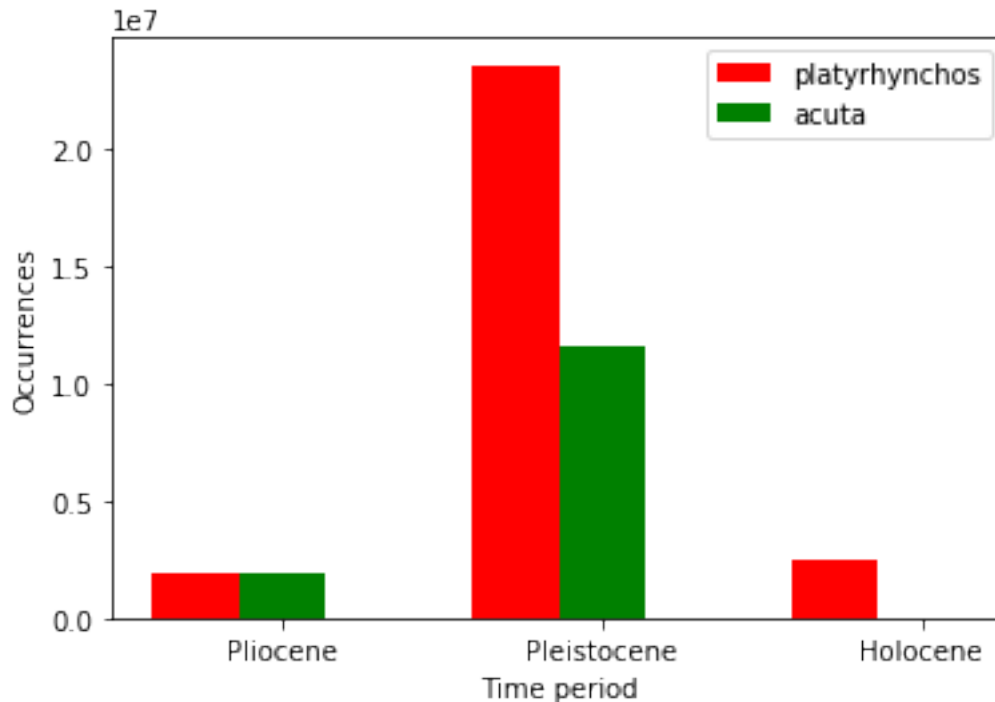
import numpy as np
import matplotlib.pyplot as plt

N = 3 # for specific time periods
ind = np.arange(N)
width = 0.27 #indicates width of the bar graph
fig = plt.figure()
ax = fig.add_subplot(111)

plat = [594522+1314658,
        1020839+1024172+1024173+1195901+
        1196100+1196603+1197044+1197459+
        1197712+1197895+1201082+1313400+
        213817+219357+219457+219504+
        219524+370509+447661+665510+709981+
        710085+740024+740518+781719+789196+
        840107+841378+915167, 1186504+1318491]
# all the occurrences listing for Anas platyrhynchos
rects1 = ax.bar(ind, plat, width, color='r')

acuta = [594523+1314658, 1083779+1196606+
        1197460+1197716+1201083+1313401+
        213766+219352+219392+219498+219518+
        447659+623927+665511+690674+841379, 0]
# all the occurrences listing for Anas acuta
rects2 = ax.bar(ind+width, acuta, width, color='g')

ax.set_ylabel('Occurrences')
ax.set_xticks(ind+width)
ax.set_xticklabels( ('Pliocene', 'Pleistocene', 'Holocene') )
# geological time scale listd from oldest to most recent
ax.set_xlabel('Time period')
ax.legend( ('platyrhynchos', 'acuta') )
plt.show()
```



This is a graph that shows the relationship between time period and occurrences of the fossil records. On the x-axis, I have the time period, in which the labels are divided into three different periods the fossil records were dated back to. Time periods are listed from oldest to youngest, in which the oldest corresponds to Pliocene and Holocene is the most recent. On the y-axis, I have the occurrences of the fossil records.

This graph shows that both *Anas acuta* and *Anas platyrhynchos* are found early in the Pliocene period, and most of the fossil records are dated back predominantly to the Pleistocene period. Looking at the bar graphs signifies that there are far more fossil records of *Anas platyrhynchos* found, especially in the Pleistocene period.

By comparing the occurrences of the fossil records found, it is easily determined that there might have been a species multiplication that started in the Pleistocene period. However, this graph still does not present to us how *Anas acuta* and *Anas platyrhynchos* might be related. Therefore, another set of graph is created using RStudio that may potentially illustrate the relationship between the two species of birds.

## 4 RStudio Graph

In order to determine further the relationship between *Anas acuta* and *Anas platyrhynchos*, I used RStudio to graph a linear-range graph. First, I used the previous dataset containing all the unique species of birds from family Anatidae to graph the linear-range plot as a reference. The graph will show where the my species-of-interest falls under and will give a general baseline.

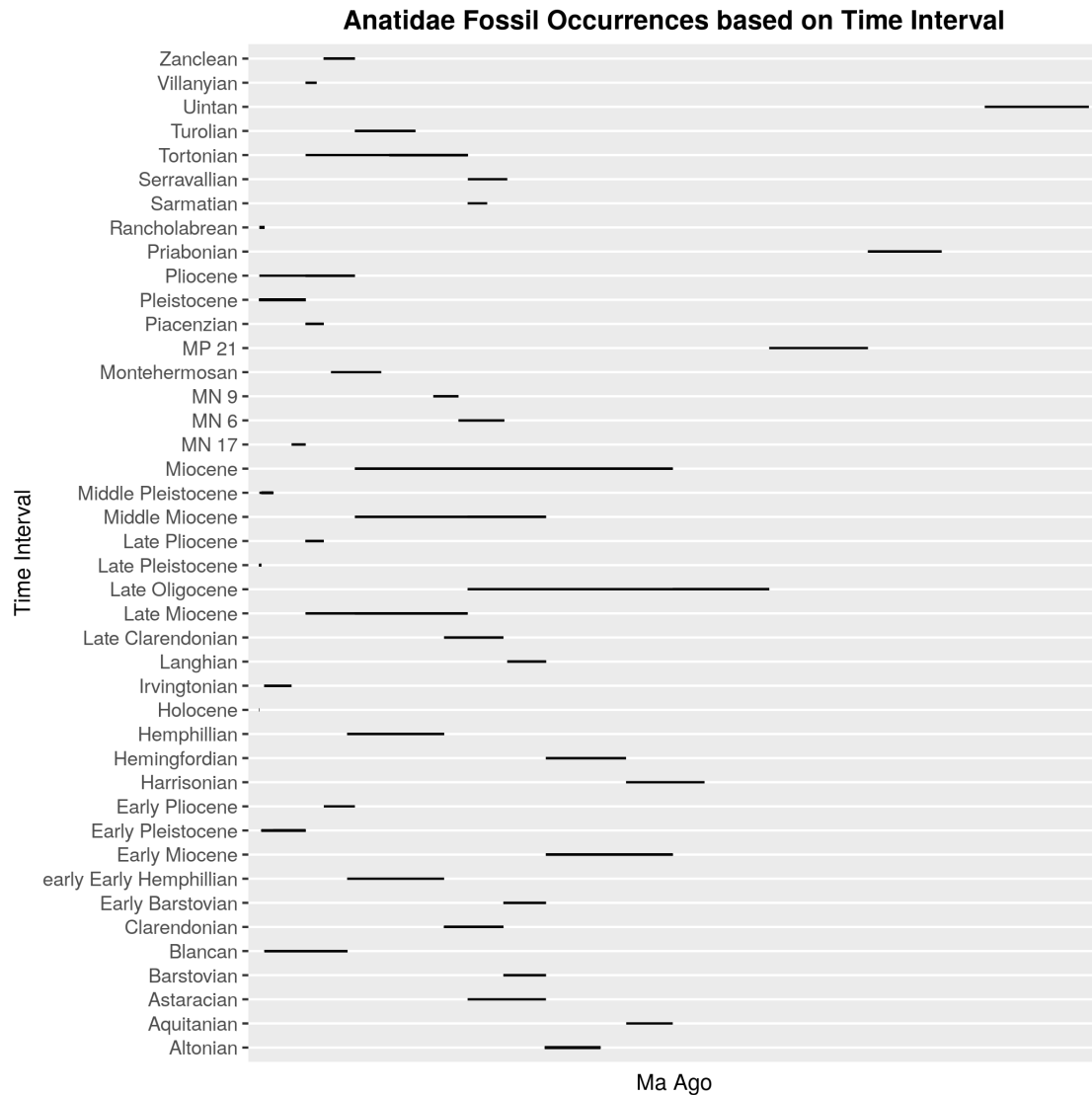
All codes for RStudio are shown in markdown, and the images of the graphs shown are imported from saved images, using the code “ggsave.”

### 4.1 RStudio Graph 1

```
library(ggplot2)
setwd("/home/eeb177-student/Desktop/eeb-177/eeb-177-final-project/final/")
p1 <- read.csv("/home/eeb177-student/Desktop/eeb-177/eeb-177-final-
project/final/anatidae_1.csv", header = F, as.is = T)
p1 <- ggplot(p1, aes(V9, ymin = V12, ymax = V11))
p1 <- p1 + geom_linerange()
p1 <- p1 + coord_flip()
p1 <- p1 + scale_y_continuous(limits=c(0, 40),expand = c(0, 0),breaks=c(0, 10, 20, 30, 40))
p1
p1 <- p1 + scale_x_discrete(name="Time Interval")+ scale_y_discrete(name="Ma Ago")
p1 <- p1 + labs(title = "Anatidae Fossil Occurrences based on Time Interval", x = "Ma ago",y
= "Time Interval") + theme(plot.title = element_text(hjust = 0.5, size=12,face = "bold"),axis.title
=element_text(size=10))
p1
```

```
In [14]: from IPython.display import Image
         Image(filename = "occurrence_species.png")
```

```
Out [14]:
```



This is the linear range plot I used to graph the occurrences of all of the unique species in family "Anatidae." The x-axis shown represents the range of the Ma. The farther the occurrences go to the right illustrates the older time period (older Earth age), and the closer the line is to the 0, or to the left, signify younger Earth age. On the y-axis, I listed the unique time interval of the fossil records.

The reason why I chose time interval as y-axis instead of the unique species, is because I was specifically interested in looking at the timeline rather than how old the species were dated back to. Plotting timeline against timeline demonstrates important information about what happened in that time period.

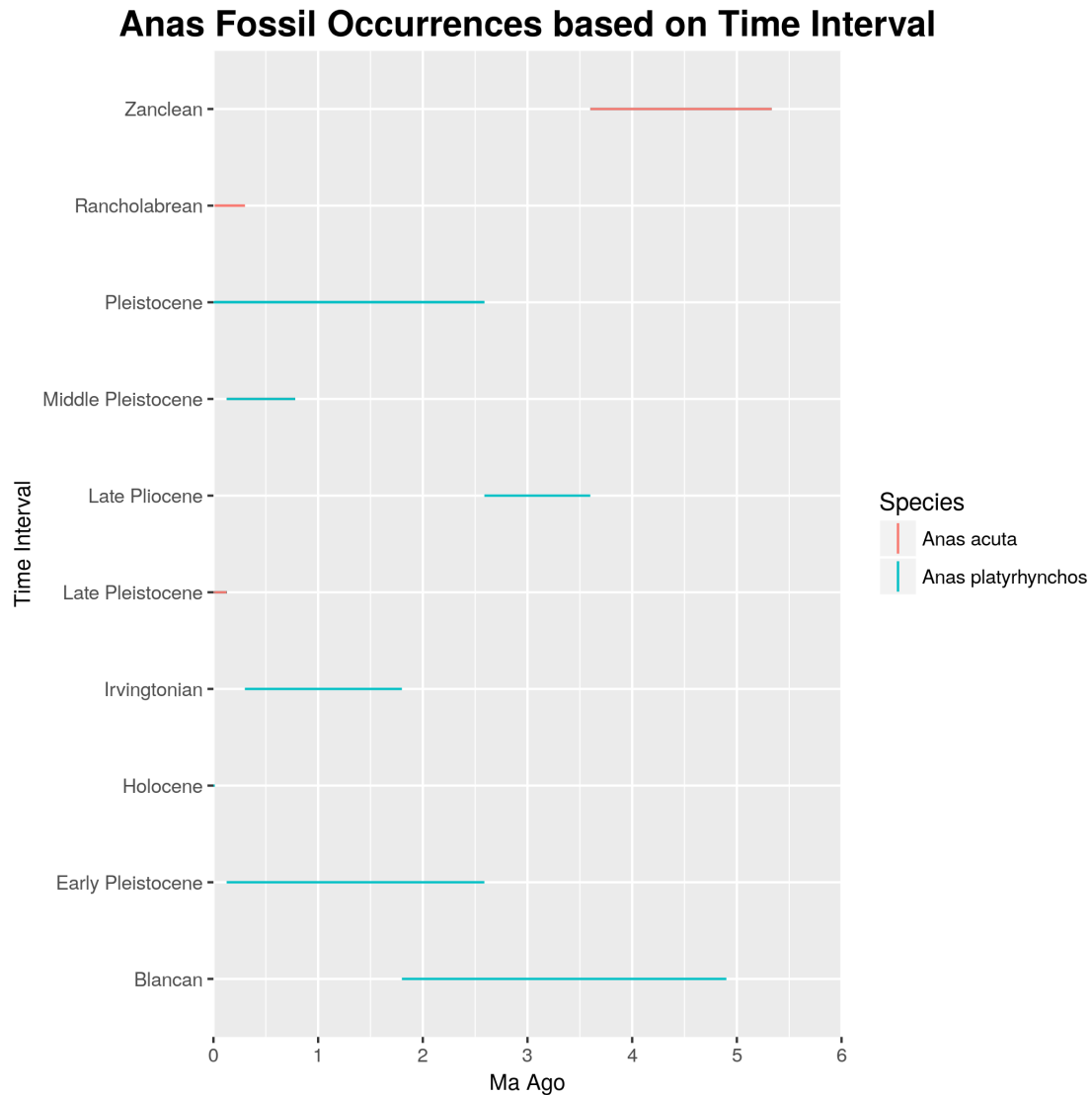
Then, the following graph shows the same plot, except that instead of plotting all of the unique species, I only took out information related to *Anas acuta* and *Anas platyrhynchos*.

## 4.2 RStudio Graph 2

```
library(ggplot2)
setwd("/home/eeb177-student/Desktop/eeb-177/eeb-177-final-project/final/")
p2 <- read.csv("/home/eeb177-student/Desktop/eeb-177/eeb-177-final-
project/final/2birds.csv", header = F, as.is = T)
p2 <- ggplot(p2, aes(V3, ymin = V5, ymax = V4, color=V2))
p2 <- p2 + geom_linerange()
p2 <- p2 + coord_flip()
p2 <- p2 + scale_y_continuous(limits=c(0, 6), expand = c(0, 0), breaks=c(0, 1, 2, 3, 4, 5, 6))
p2
p2 + scale_x_discrete(name="Time Interval") + scale_y_discrete(name="Range") +
scale_color_discrete(name="species")
p2 <- p2 + labs(title = "Anas Fossil Occurrences based on Time Interval", x = "Ma ago", y =
"Time Interval", color = "Species") + theme(plot.title = element_text(hjust = 0.5, size=16, face =
"bold"), axis.title = element_text(size=10))
p2
```

```
In [21]: from IPython.display import Image
Image(filename = "two-species.png")
```

Out[21]:



This is the linear range plot I used to graph the occurrences of *Anas acuta* and *Anas platyrhynchos*. The x-axis shown represents the range of the Ma. The farther the occurrences go to the right illustrates the older time period (older Earth age), and the closer the line is to the 0, or to the left, signify younger Earth age. On the y-axis, I listed the unique time interval of the fossil records.

This graph shows that the earliest fossil records of *Anas acuta* was dated back to the period of Zanclean, around 5.5 Ma ago. The earliest fossil record of *Anas platyrhynchos* was dated back to the period of Blancan, around 5 Ma ago. This signifies that *Anas platyrhynchos* diverged from *Anas acuta* around 5 Ma ago, sharing the same ancestors.

## 5 Conclusion

As you can see, *Anas acuta* and *Anas platyrhynchos* are related to one another by *Anas platyrhynchos* diverging from *Anas acuta*. The Pdbb dataset provided with extensive information about the fossil records, allowing the determination of the relationship between the two species.

Because these two bird species are related to one another, the genetic relatedness made them possible to carry the same influenza strain. There may be some conserved sequence in their genome that the influenza strain strongly binds to, therefore allowing the bird native to North America to carry Asian-origin influenza strain.

To go beyond the research, it may be interesting to research which DNA sequence the influenza strain binds to, and to observe how conserved the genetic information is across the two species, *Anas acuta* and *Anas platyrhynchos*.

## 6 Reference

Sax DF, Gaines SD, and Brown JH. Species invasions exceed extinctions on islands worldwide: a comparative study of plants and birds. 2002. *Am Nat.* 160(6):766-83

Kevin P. Johnson and Michael D. Sorenson. Phylogeny and Biogeography of Dabbling Ducks (Genus: *Anas*): A Comparison of Molecular and Morphological Evidence. 1999. *The Auk.* 116(3): 792-805

Anson V. Koehler, John M. Pearce, Pearce, Paul L. Flint, J. Christian Franson and and Hon S. IP. Genetic evidence of intercontinental movement of avian influenza in a migratory bird: the northern pintail (*Anas acuta*). 2008. *Molecular Ecology.* 17(21): 4754–4762