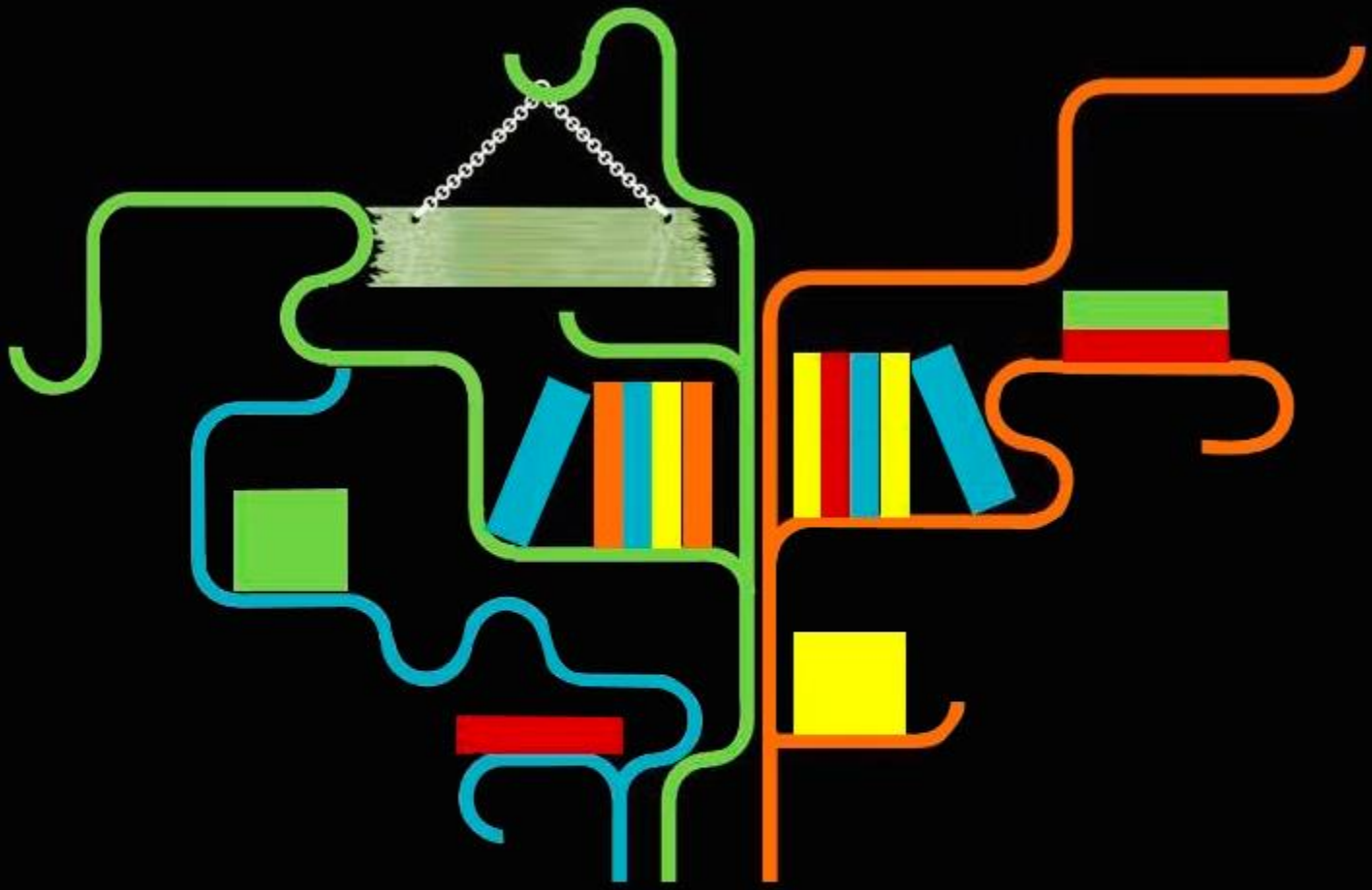
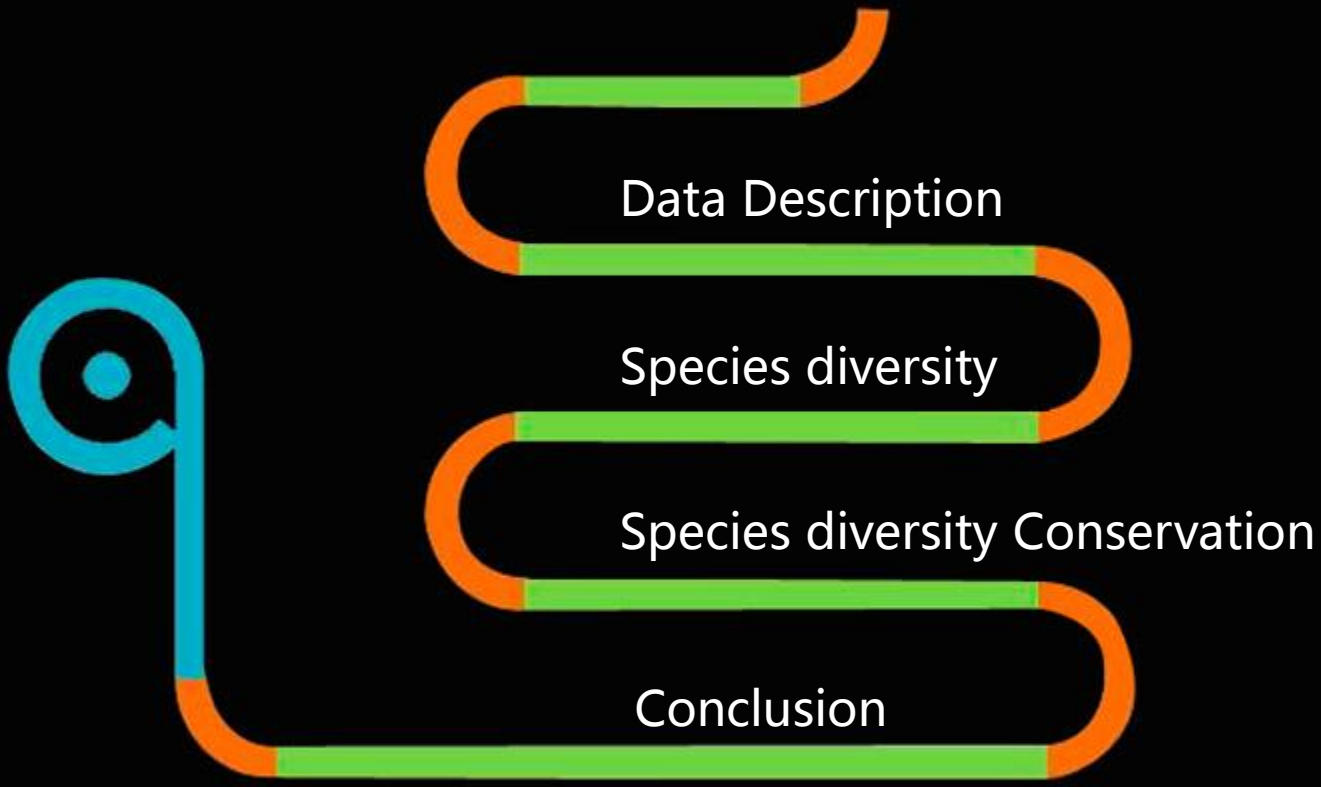


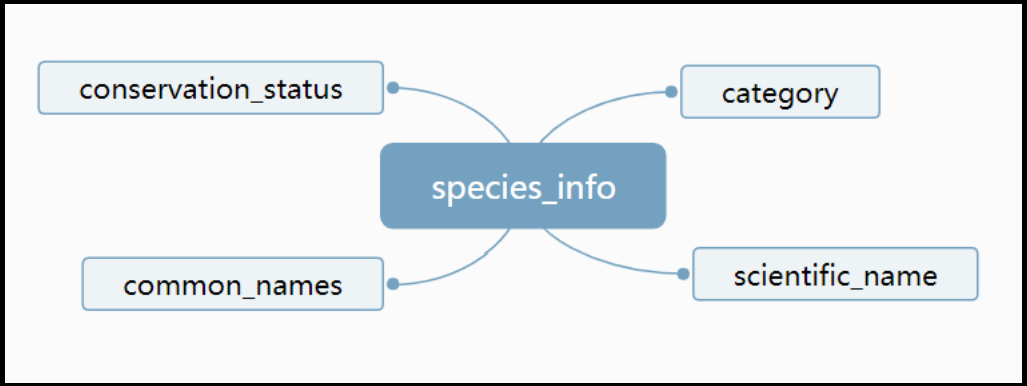
biodiversity

By Yijia Ling

CONTENT


Data Description
Species diversity
Species diversity Conservation
Conclusion

Data Description



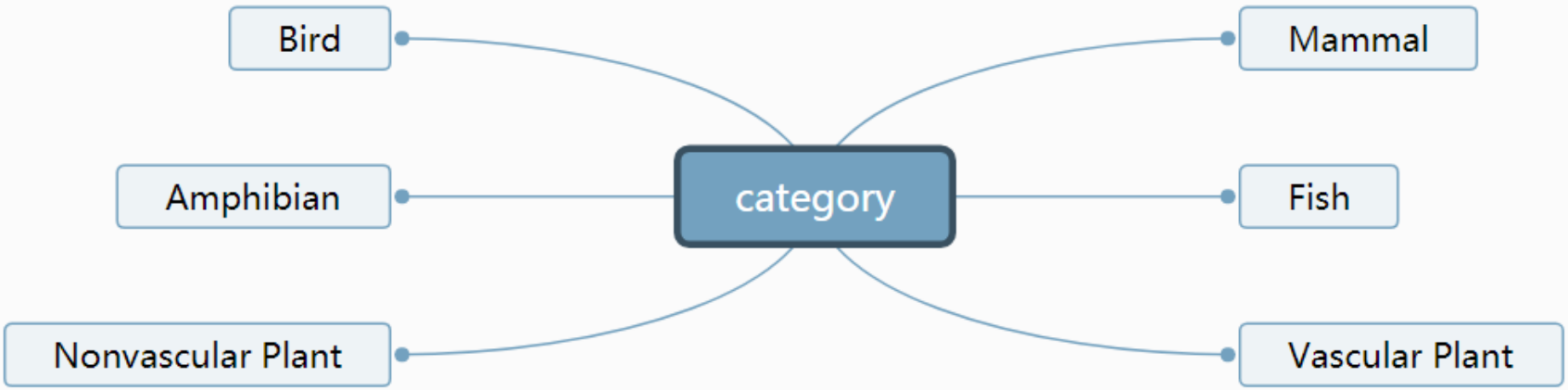
species.scientific_name.nunique()
>>>5541 !
So many different species in our National Parks

```
species.head(10)
```

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	No Intervention	False	False
1	Mammal	Bos bison	American Bison, Bison	No Intervention	False	False
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	No Intervention	False	False
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
4	Mammal	Cervus elaphus	Wapiti Or Elk	No Intervention	False	False
5	Mammal	Odocoileus virginianus	White-Tailed Deer	No Intervention	False	False
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	No Intervention	False	False
7	Mammal	Canis latrans	Coyote	Species of Concern	True	False
8	Mammal	Canis lupus	Gray Wolf	Endangered	True	False
9	Mammal	Canis rufus	Red Wolf	Endangered	True	False

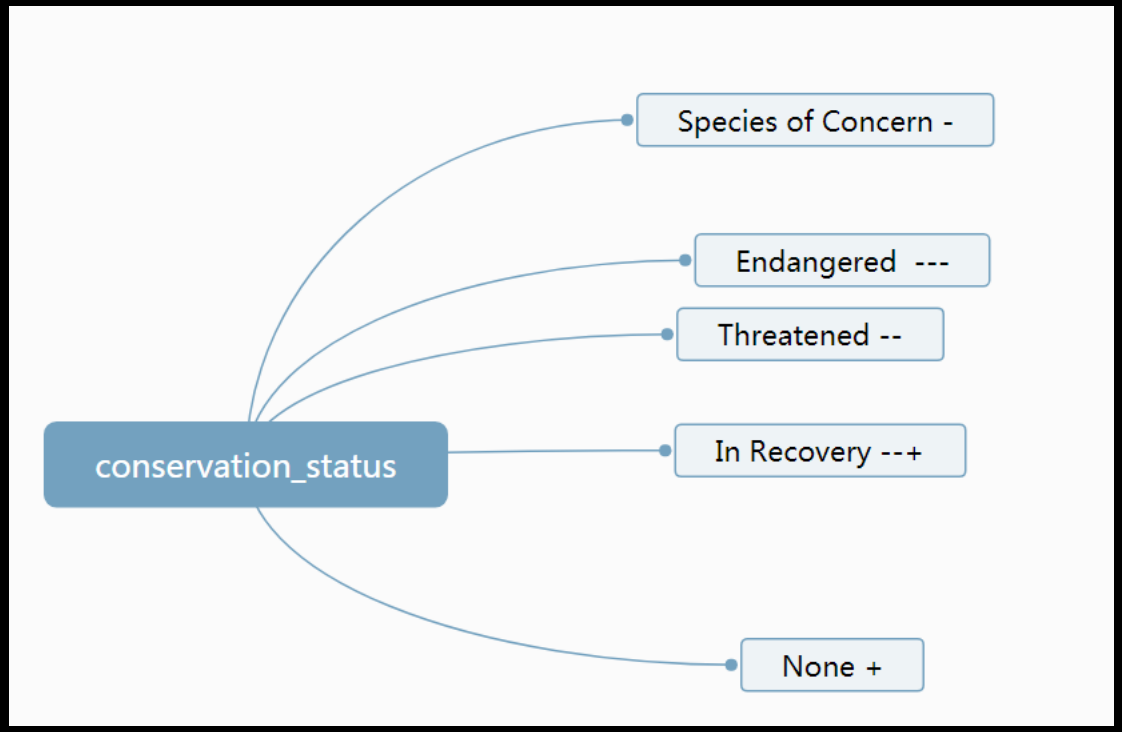
Data Description

```
species.category.unique()
>>>array(['Mammal', 'Bird', 'Reptile',
'Amphibian', 'Fish', 'Vascular Plant',
       'Nonvascular Plant'], dtype=object)
```



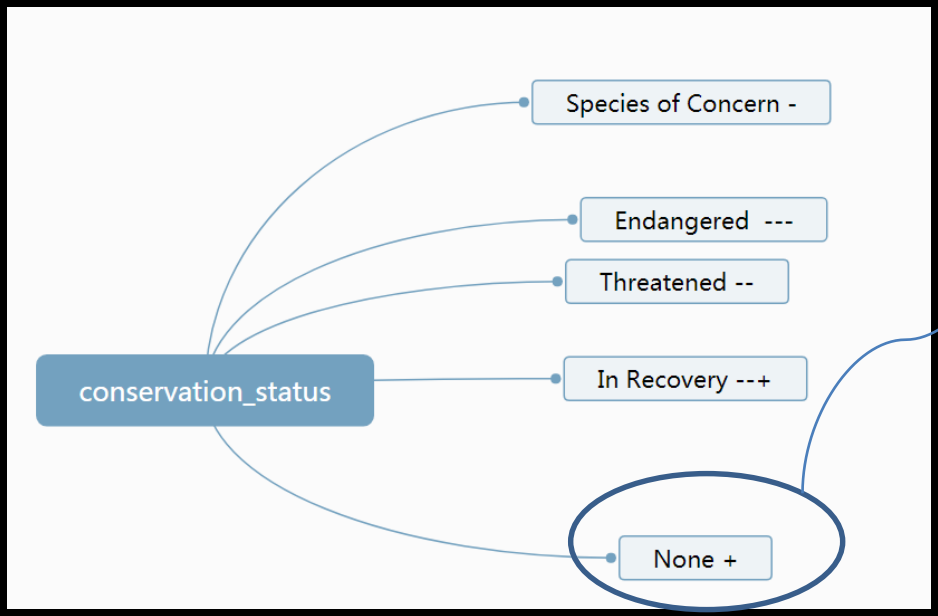
Data Description

```
species.conservation_status.unique()
>>>array([nan, 'Species of Concern',
'Endangered', 'Threatened',
       'In Recovery'], dtype=object)
```



So we know that some of them are in protection others are not!

1 Species diversity



1 We replace the nan value to ‘No Intervention’ (a new status)

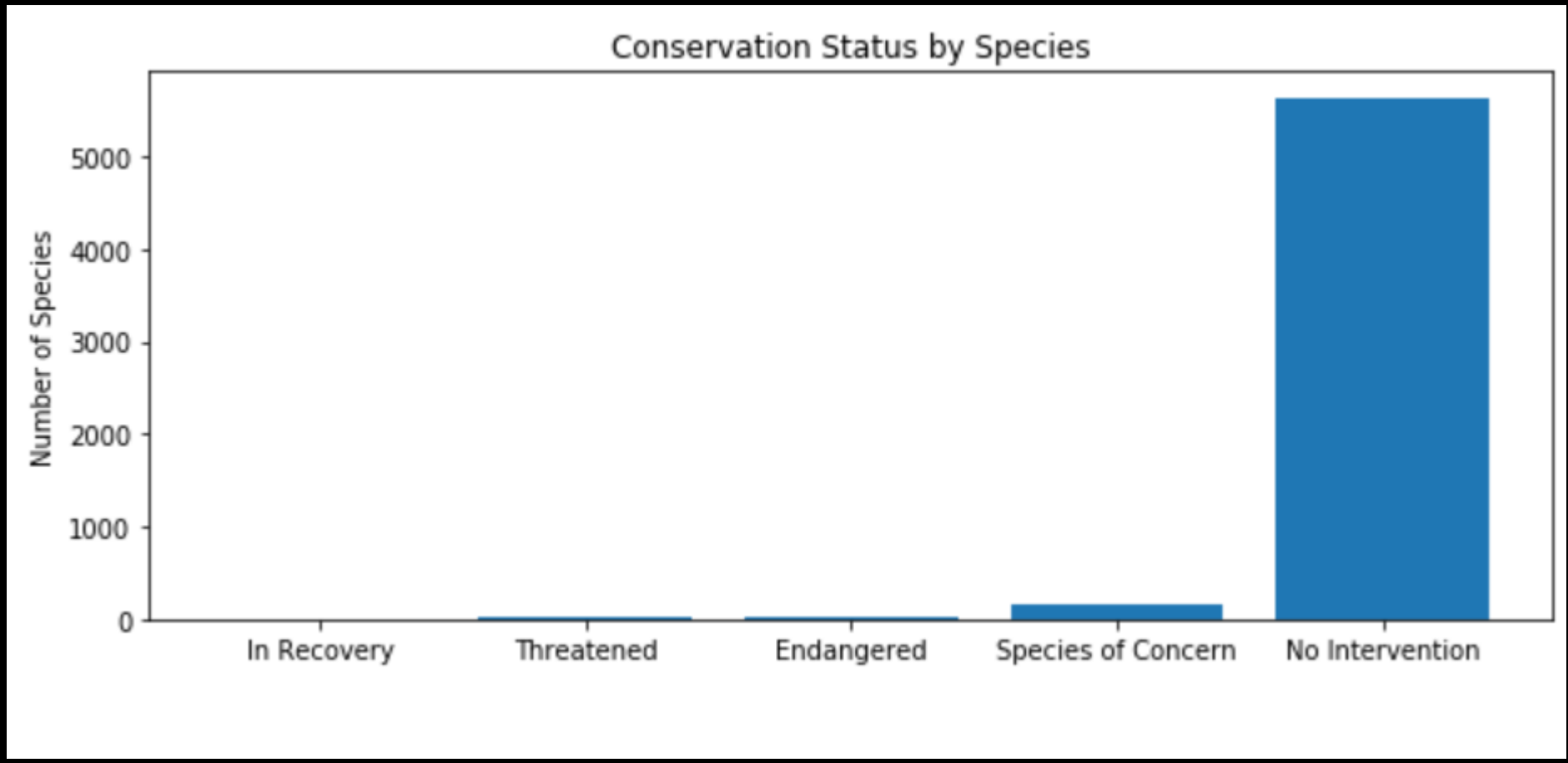
2 most of species are in normal number

```
In [21]: species.groupby('conservation_status').scientific_name.nunique().reset_index()
```

Out[21]:

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

1 Species diversity



By using plot function in matplotlib, we've got bar char of different species' conservation status. How to protect these endangered and threatened species is important.

2 Species diversity Conservation – Part A

If there are some category of species have higher possibility to extinction?
Let's create a new column in species called is_protected, which is True if conservation_status is not equal to No Intervention, and False otherwise.

```
category_counts = species.groupby(['category',
'is_protected']).scientific_name.nunique().reset_index()
category_counts.head()
```

	category	is_protected	scientific_name
0	Amphibian	False	72
1	Amphibian	True	7
2	Bird	False	413
3	Bird	True	75
4	Fish	False	115

2 Species diversity Conservation – Part A

	category	is_protected	scientific_name
0	Amphibian	False	72
1	Amphibian	True	7
2	Bird	False	413
3	Bird	True	75
4	Fish	False	115

We want a more clear table!
So let us change the original table to a pivot table.

It looks like species in category Mammal are more likely to be endangered than species in Bird?

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

2 Species diversity Conservation – Part A

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Mammal & Bird

Is the data numerical or categorical?
Categorical

chi squared test

```
from scipy.stats import chi2_contingency
contingency = [[30, 146],
               [75, 413]]
chi2_contingency(contingency)
>>>pval=0.68
```

We **don’t** reject H0:There's no significant difference between the datasets

2 Species diversity Conservation – Part A

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Mammal & Reptile

Is the data numerical or categorical?
Categorical

chi squared test

```
from scipy.stats import chi2_contingency
contingency = [[30, 146],
               [5, 73]]
chi2_contingency(contingency)
>>>pval=0.04
```

We reject H0:There's no significant difference between the datasets

2 Species diversity Conservation – Part B

Our scientists got samples to test the Species diversity in National Park!

observations.csv

```
observations = pd.read_csv('observations.csv')
observations.head()
```

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

2 Species diversity Conservation – Part B

Some scientists are studying the number of **sheep** sightings at different national parks.

```
species['is_sheep'] = species.common_names.apply(lambda x: 'Sheep' in x)
species.head()
```

Now we got the data that only contain sheep species !

```
species[species.is_sheep]
```

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
1139	Vascular Plant	Rumex acetosella	Sheep Sorrel, Sheep Sorrell	No Intervention	False	True
2233	Vascular Plant	Festuca filiformis	Fineleaf Sheep Fescue	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3758	Vascular Plant	Rumex acetosella	Common Sheep Sorrel, Field Sorrel, Red Sorrel,...	No Intervention	False	True
3761	Vascular Plant	Rumex paucifolius	Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock	No Intervention	False	True
4091	Vascular Plant	Carex illota	Sheep Sedge, Smallhead Sedge	No Intervention	False	True
4383	Vascular Plant	Potentilla ovina var. ovina	Sheep Cinquefoil	No Intervention	False	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

2 Species diversity Conservation – Part B

Some scientists are studying the number of **sheep** sightings at different national parks.

```
species['is_sheep'] =
species.common_names.apply(la
mbda x: 'Sheep' in x)
species.head()
```

Now we got the data that only contain sheep species !

```
sheep_species =
species[(species.is_sheep) &
(species.category == 'Mammal')]
sheep_species
```

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

2 Species diversity Conservation – Part B

To check the difference between sample and population, we merge two table together.

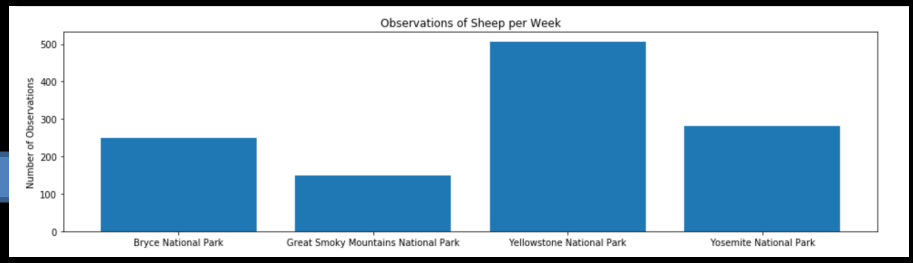
	category	scientific_name	common_names	conservation_status	is_protected	is_sheep	park_name	observations
0	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yosemite National Park	126
1	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Great Smoky Mountains National Park	76
2	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Bryce National Park	119
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yellowstone National Park	221
4	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Yellowstone National Park	219
5	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Bryce National Park	109
6	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Yosemite National Park	117
7	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Great Smoky Mountains National Park	48
8	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Yellowstone National Park	67
9	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Yosemite National Park	39
10	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Bryce National Park	22
11	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True	Great Smoky Mountains National Park	25

2 Species diversity Conservation – Part B

How many total sheep observations (across all three species) were made at each national park?

```
obs_by_park =
sheep_observations.groupby('park_name').o
bservations.sum().reset_index()
obs_by_park
```

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282



2 Species diversity Conservation – Part B

How many total sheep observations (across all three species) were made at each national park?

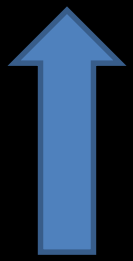
	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

2 Species diversity Conservation – Part C

Sheep foot and mouth disease happens at Bryce National Park！
（15%）
Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park.

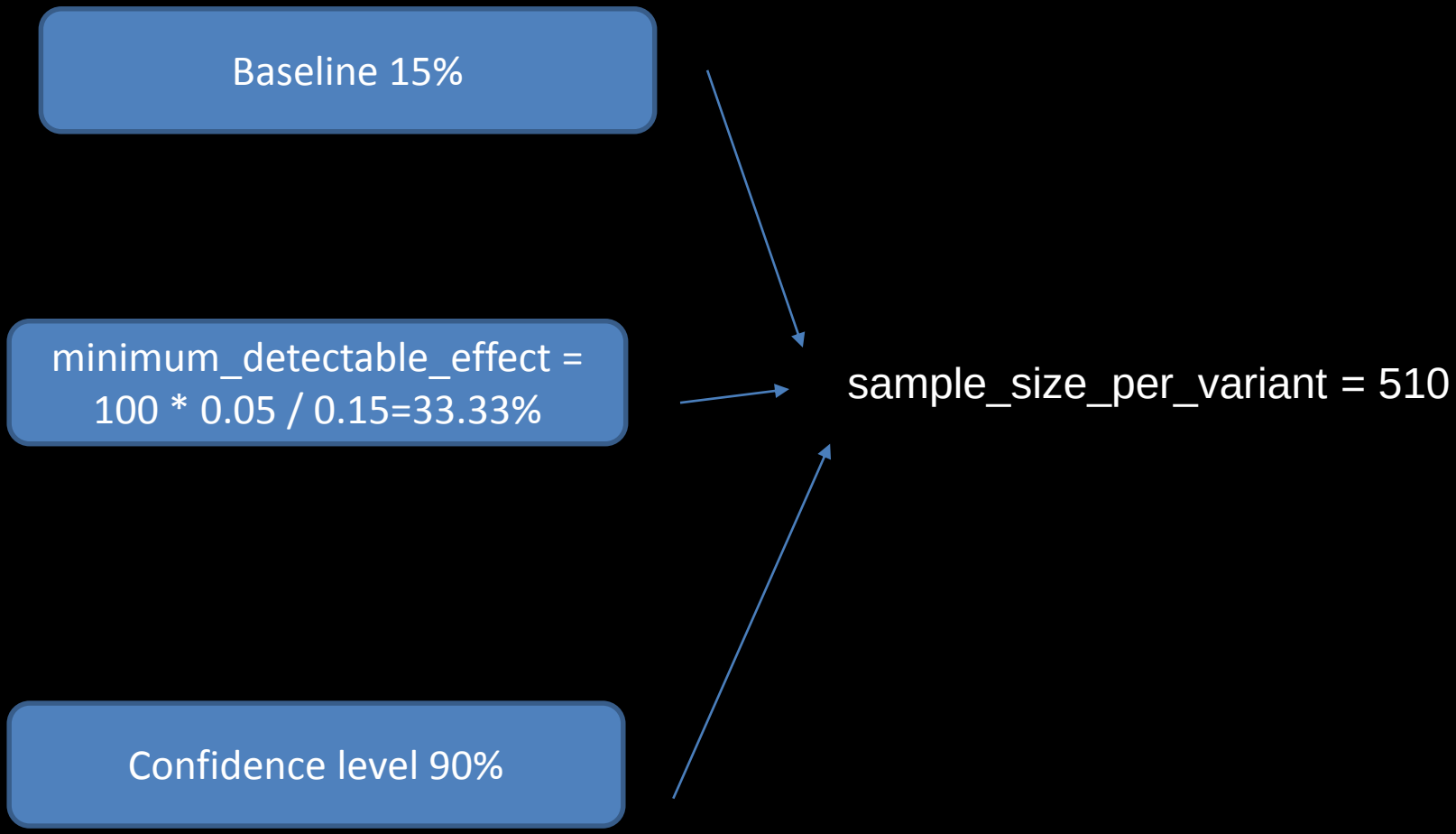
Baseline 15%

minimum_detectable_effect =
100 * 0.05 / 0.15=33.33%



They want to be able to detect reductions of at least 5 percentage point. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

2 Species diversity Conservation – Part C



2 Species diversity Conservation – Part C

How many weeks would you need to observe sheep at Bryce National Park in order to observe enough sheep? How many weeks would you need to observe at Yellowstone National Park to observe enough sheep?

sample_size_per_variant = 510

```
bryce = 510 / 250.
yellowstone = 510 / 507.

# Approximately 2 weeks at Bryce and 1 week at Yellowstone.
```