PD.READ_CSV('SPECIES_INFO.CSV')

Species_info.csv is like a dirty room.

- Large volume of data
- Repeated values
- Poor categorization

But thankfully we have .groupby, ,count(), .unique(), .fillna(), .sort_values() to clean up our mess!

These are various functions that help solve the issues listed.

And now we can find what we were looking for!

- Significance Levels
- Sample Sizes
- Solutions!

CHI2_CONTINGENCY(CONTINGENCY)

H_o: There is not a significant diff. b/w the endangerment of Birds/Reptiles v. Mammals

H_1: There is a significant diff. b/w the endangerment of Birds/Reptiles v. Mammals

VS.

=

H_o

(P value =.53)

VS.

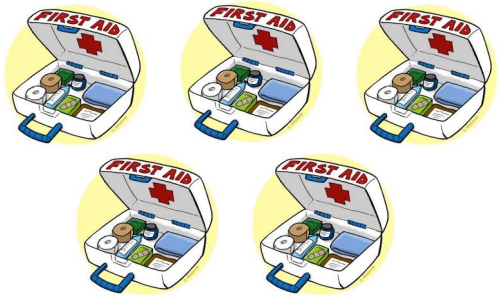
=

H_1

(P value =.05)

(CONTINGENCY) VS. (CONTINGENCY2)

Based on the finding that some species are significantly more at risk of becoming endangered than others, conservations should develop a strategy in which they allocate greater resources to the more at risk prior to those that are less at risk.



SAMPLE_SIZE_PER_VARIATION = 510

"Our scientist know that 15% of sheep at Bryce National Park have foot and mouth disease."

Baseline Conversion Rate

15%

"They want to be able to detect reductions of at least 5 percentage points."

Minimum Detectable Effect

100 * (0.5/.15)=

33%

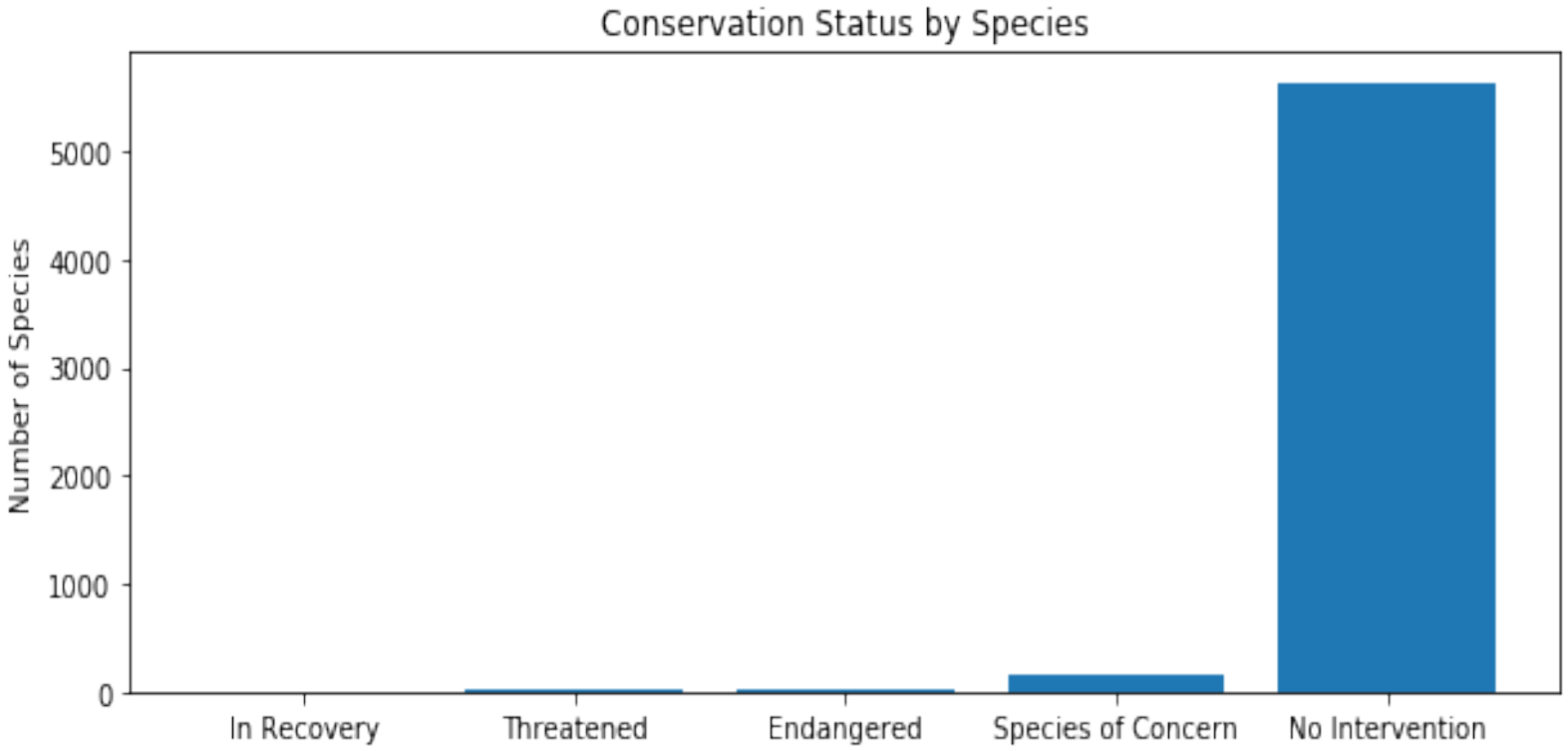
"Use the default level of significance (90%)."

Level of Significance

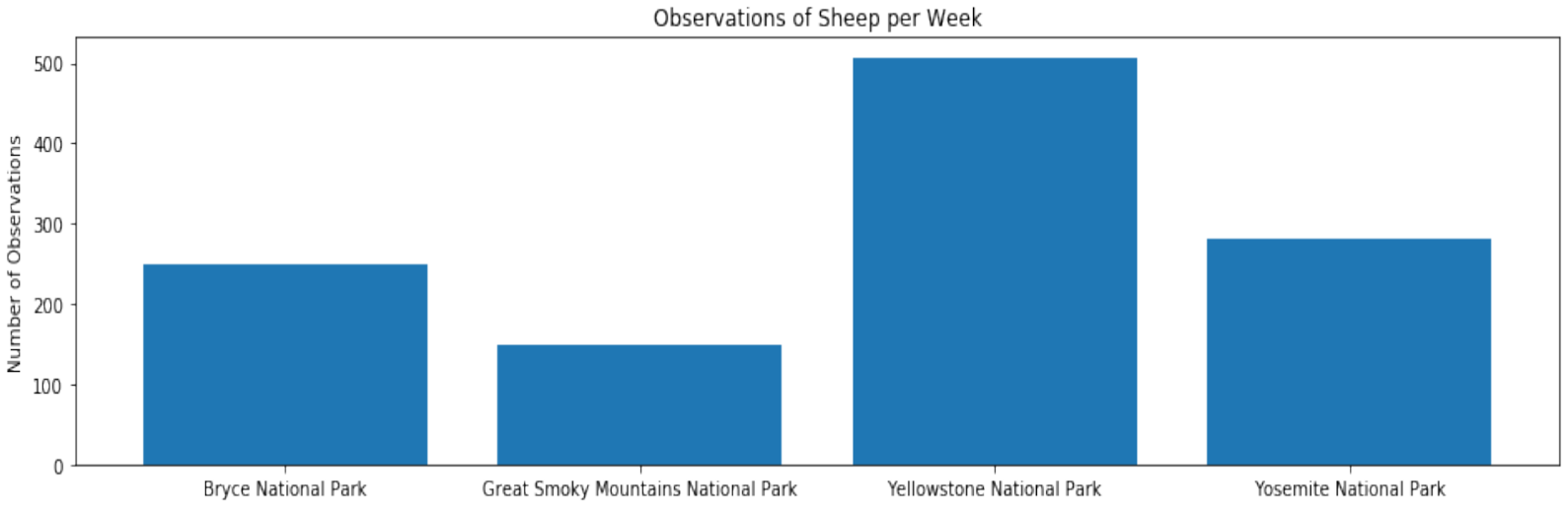
90%

OPTIMIZELY = **510**

APPENDIX 1


Conservation Status by Species
Number of Species
5000
4000
3000
2000
1000
0
In Recovery
Threatened
Endangered
Species of Concern
No Intervention

APPENDIX 2


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park