

# Modeling the Spread of Invasive Species

Jonathan Zhu  
Programming for Scientists 02-601  
December 4, 2023

# Background

- Invasive species: any organism found outside its native range
- Reducing spread: one key to stopping their impact
- Modeling spread: helps to reduce spread
- **How do we best model spread?**



Two common invasive species in North America, the lionfish and the English Ivy. Both images courtesy of wikipedia.org

# Paper Inspiration: Mainali et al. (2015)

- Use ML approaches to model spread of *Parthenium* weed
- Based on climatic and anthropogenic variables
- **Project: Replicate the model(s) used in this paper**
  - (with a few twists)
  - Not as many variables
  - Not as many data sources



The subject of interest for Mainali et al., the *Parthenium* weed, aka feverfew. Image from [tribune.com.pk](http://tribune.com.pk)

# More Accessible Modeling

## Traditional Approaches

- Specialized packages (biomod2, raster, etc)
- Require old file formats
- Hard to learn
- Skills generally not applicable to other areas
- **More reliable**

## Simpler Approach

- **Widely used packages (tidymodels suite)**
- **Utilize common file formats**
- **Easier to learn**
- **Skills more applicable to wider range of topics**
- Less reliable

# Organisms



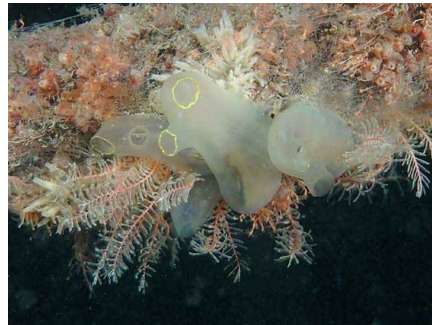
Zebra Mussel *Dreissena Polymorpha*  
Image from texasmonthly.com



*Caulerpa taxifolia*  
Image from cisr.ucr.edu



Spotted Lanternfly *Lycorma delicatula*  
Image from wikipedia.org



*Ciona intestinalis*  
Image from invasions.si.edu



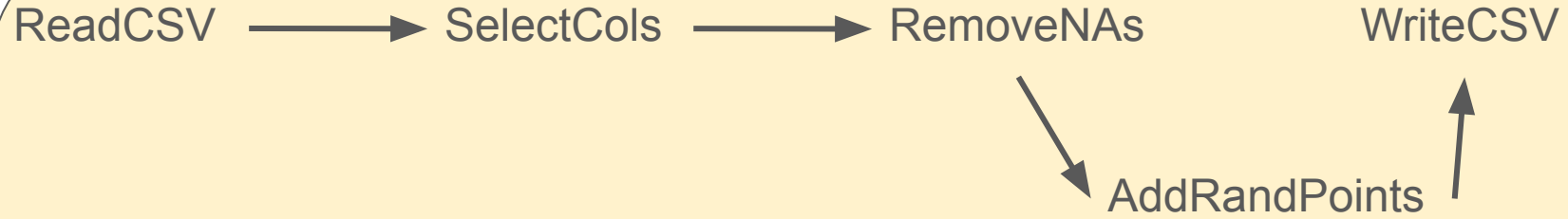
*Craspedacusta sowerbii*  
Image from lifeinfreshwater.net

# Golang: Preprocessing

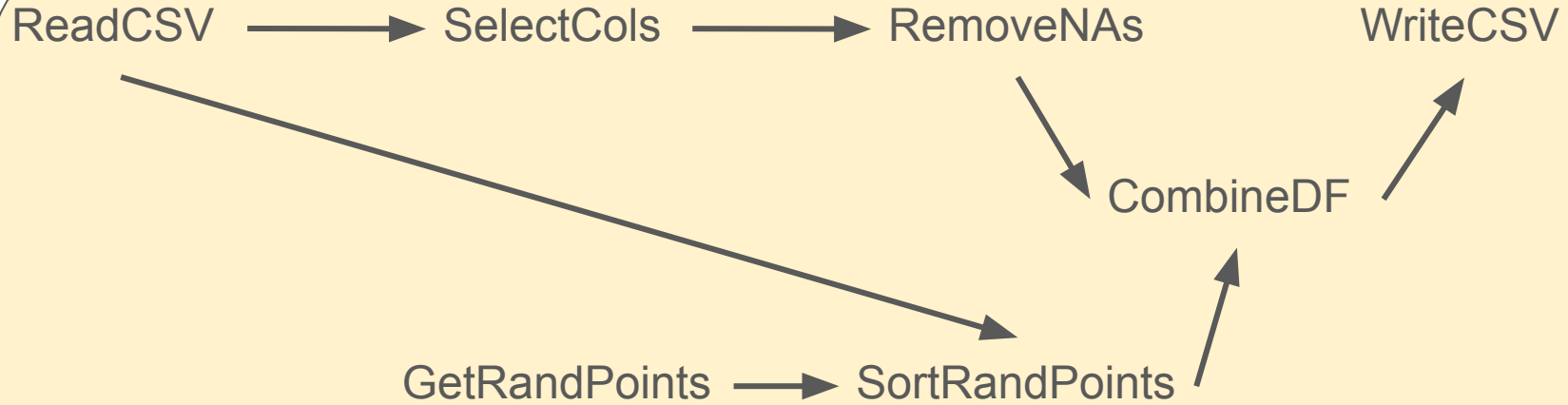
```
graph LR; ReadCSV --> SelectCols; SelectCols --> RemoveNAs; RemoveNAs --> WriteCSV;
```

ReadCSV → SelectCols → RemoveNAs → WriteCSV

# Golang: Preprocessing

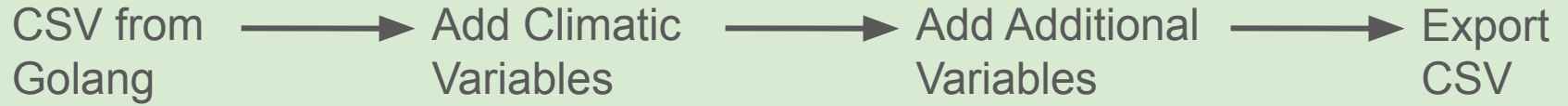


# Golang: Preprocessing





## R: Additional Processing



# Golang: Numerical EDA

GetSummaryStats

Mean  
Variance  
StdDev

FiveNumSummary

Min  
Quart1  
Median  
Quart3  
Max

GetCorrMat

Correlation

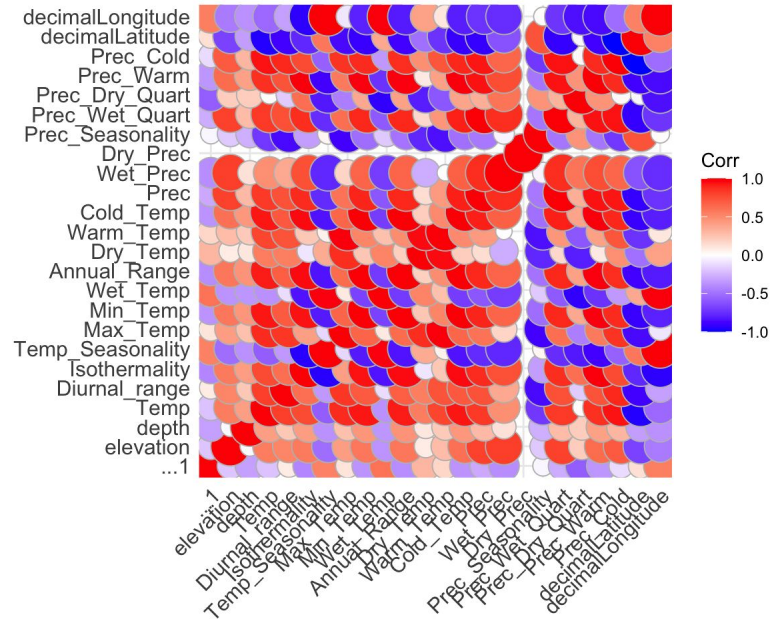
# Golang: Numerical EDA

GetSummaryStats

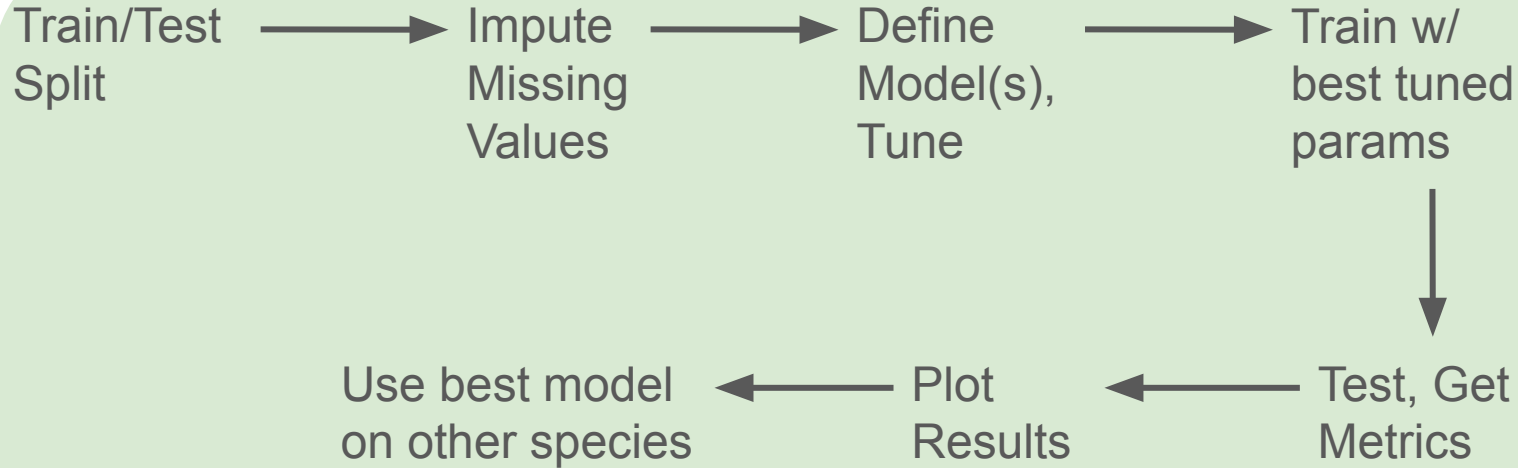
Mean  
Variance  
StdDev

FiveNumSummary

Min  
Quart1  
Median  
Quart3  
Max

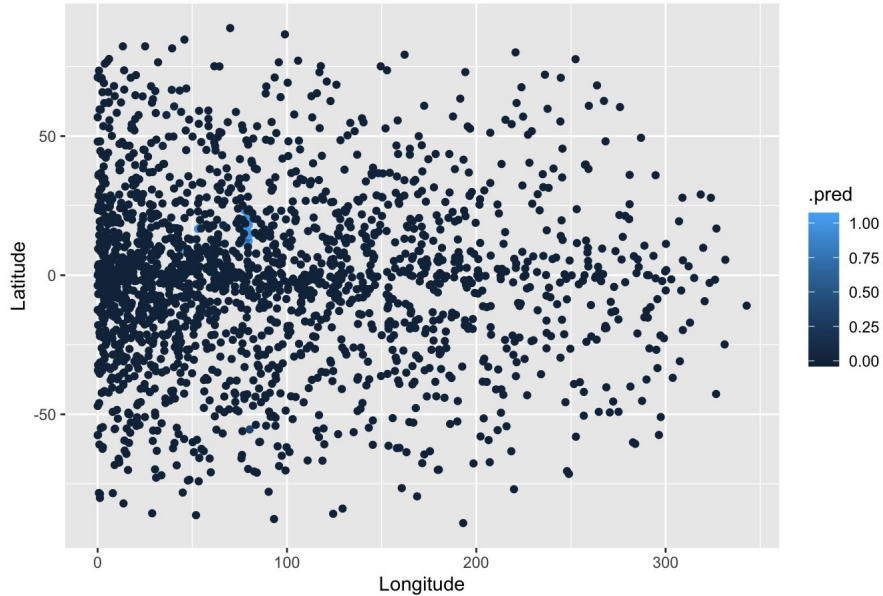


# R: Machine Learning Component (Tidymodels)

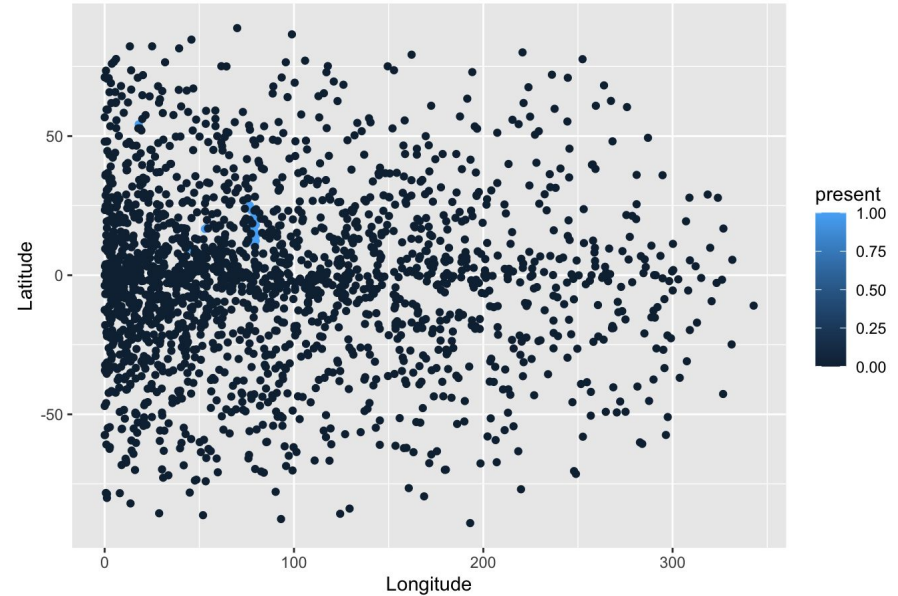


# R: Machine Learning Component

Predicted Contour of Invasive Species Spread Probability

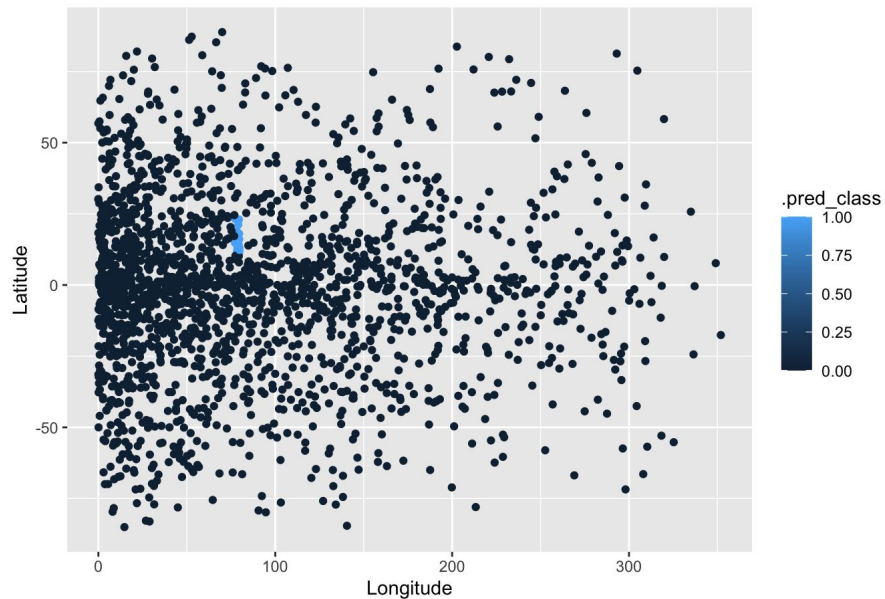


Actual Contour of Invasive Species Spread Probability

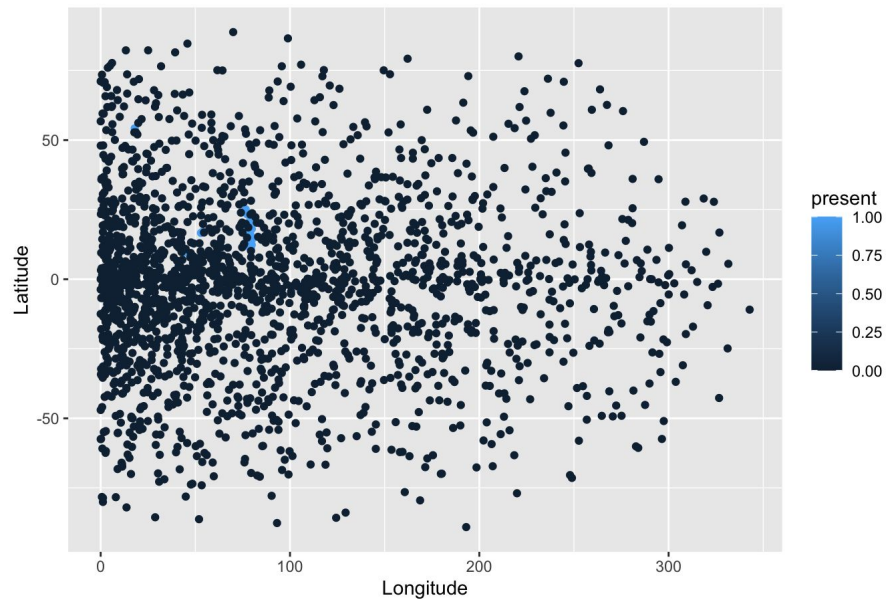


# R: Machine Learning Component

Predicted Contour of Invasive Species Spread Probability

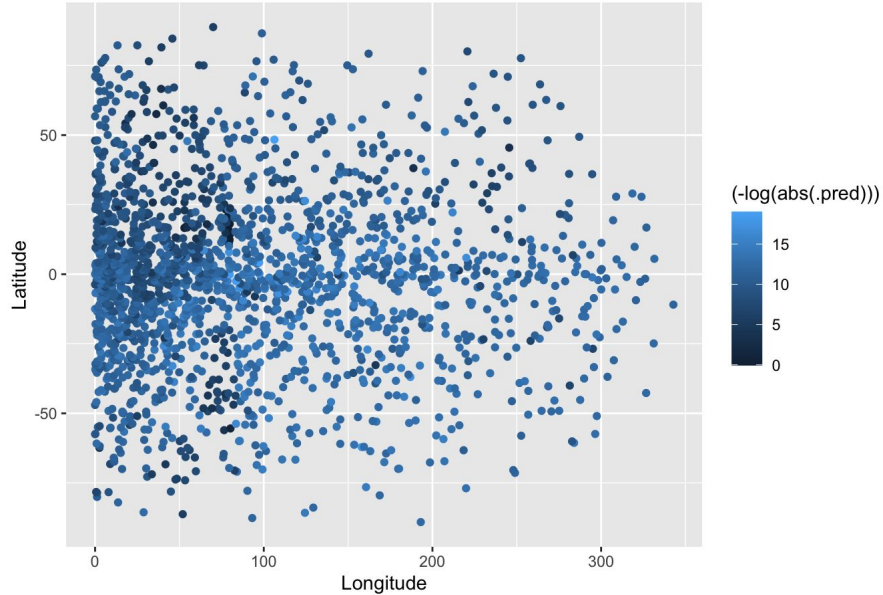


Actual Contour of Invasive Species Spread Probability

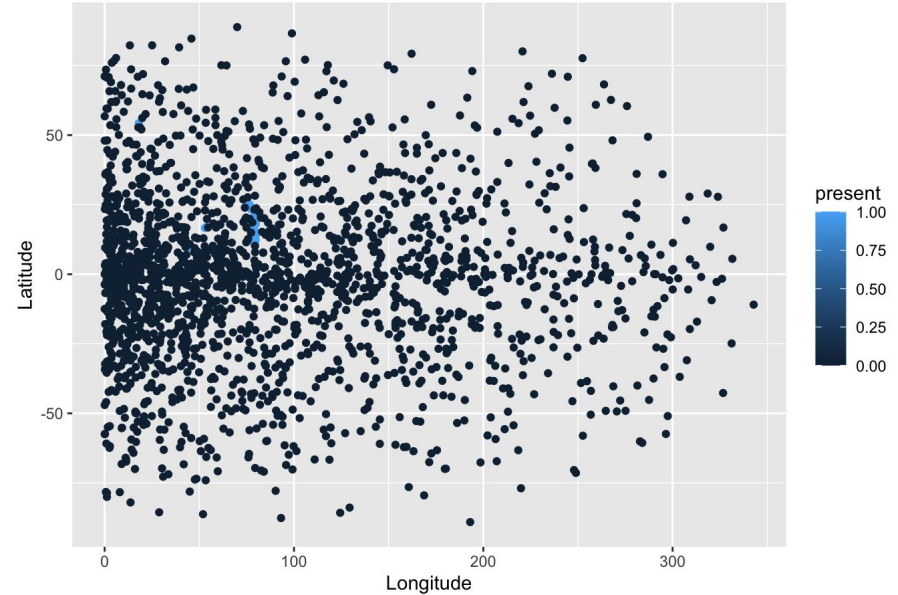


# R: Machine Learning Component

Predicted Contour of Invasive Species Spread Probability



Actual Contour of Invasive Species Spread Probability



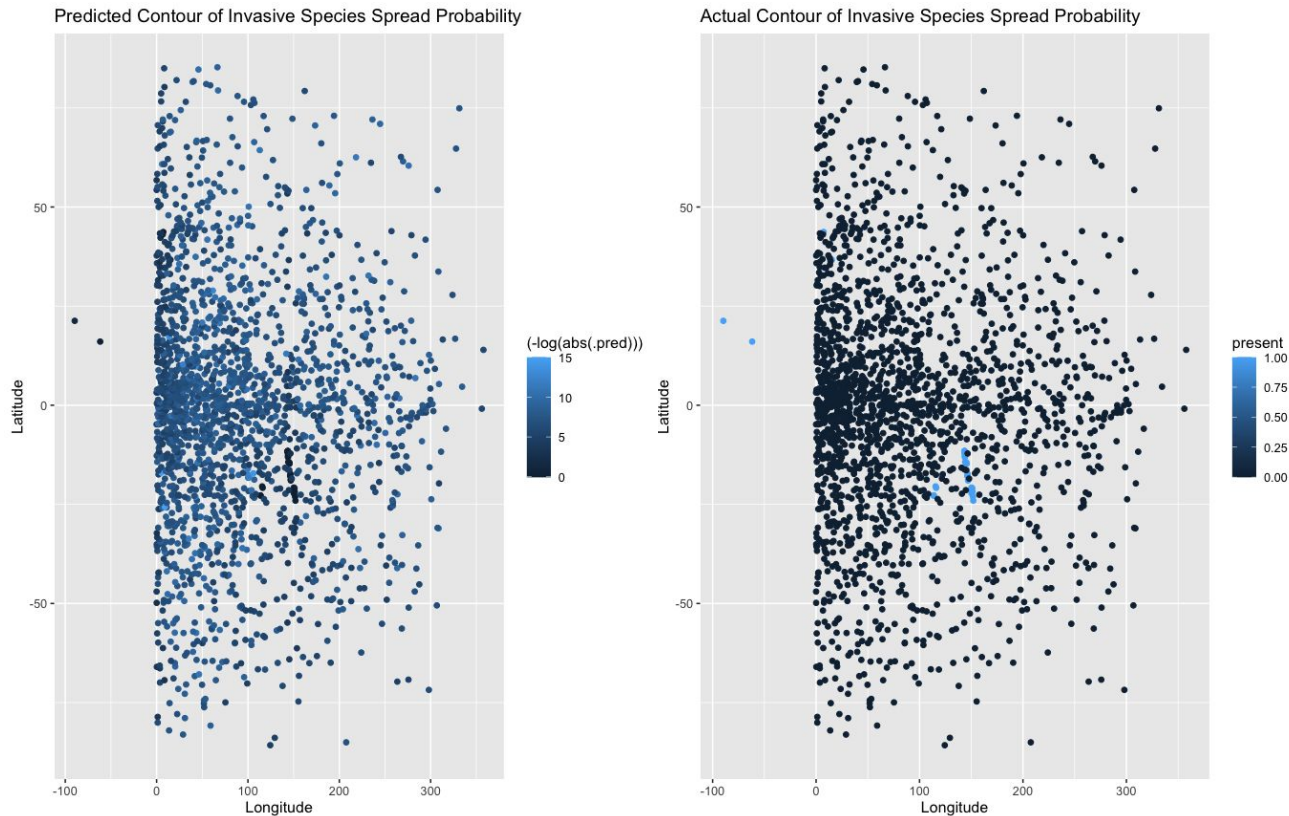


# Our Other Species

*Caulerpa taxifolia*

RMSE: 0.0219

$R^2$ : 0.976



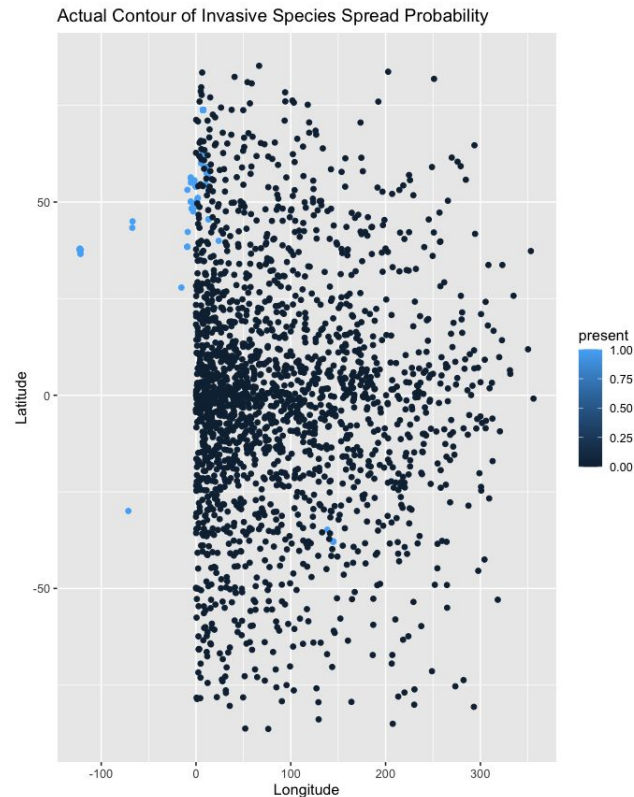
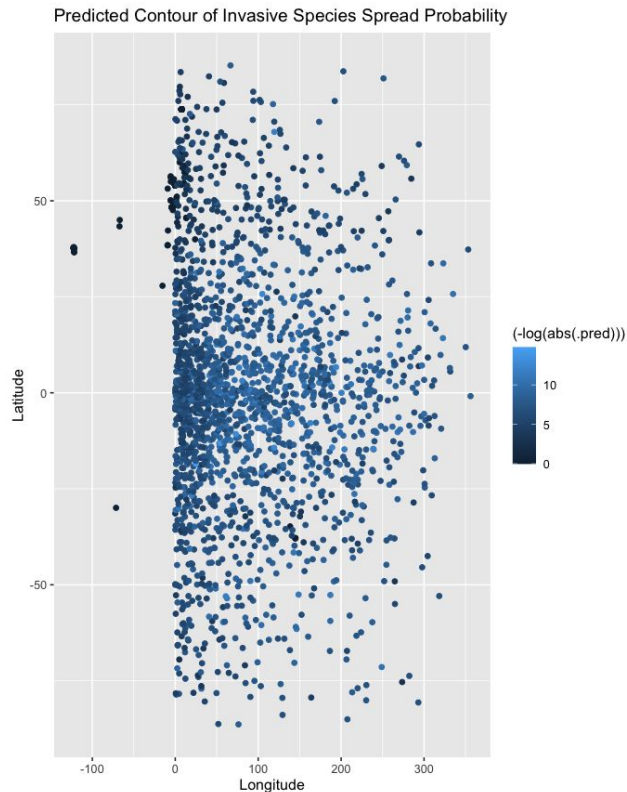


# Our Other Species

*Ciona intestinalis*

RMSE: 0.119

$R^2$ : 0.999

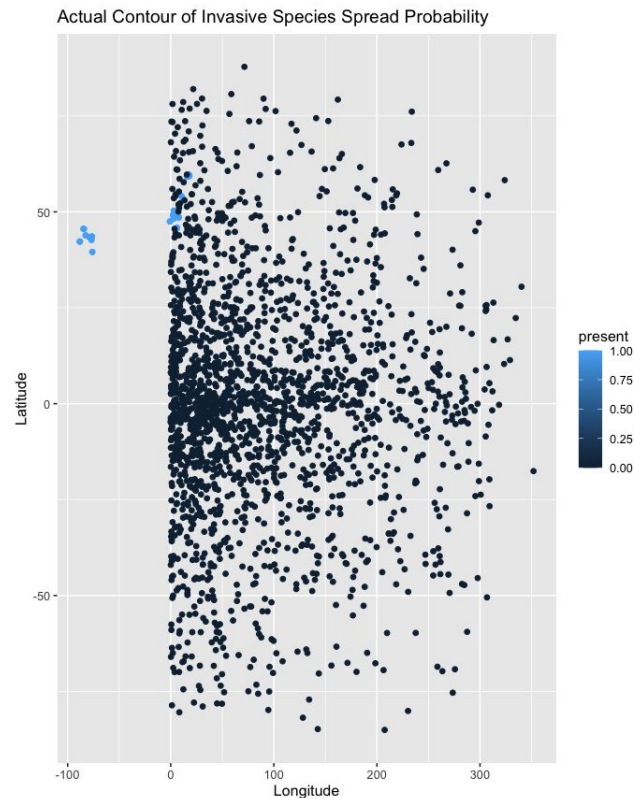
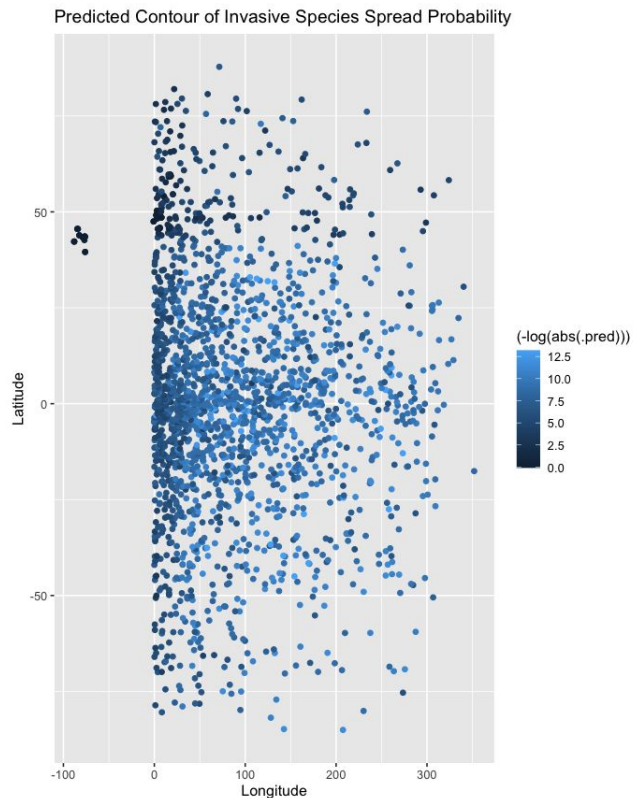


# Our Other Species

*Dreissena polymorpha*

RMSE: 0.0236

$R^2$ : 0.984

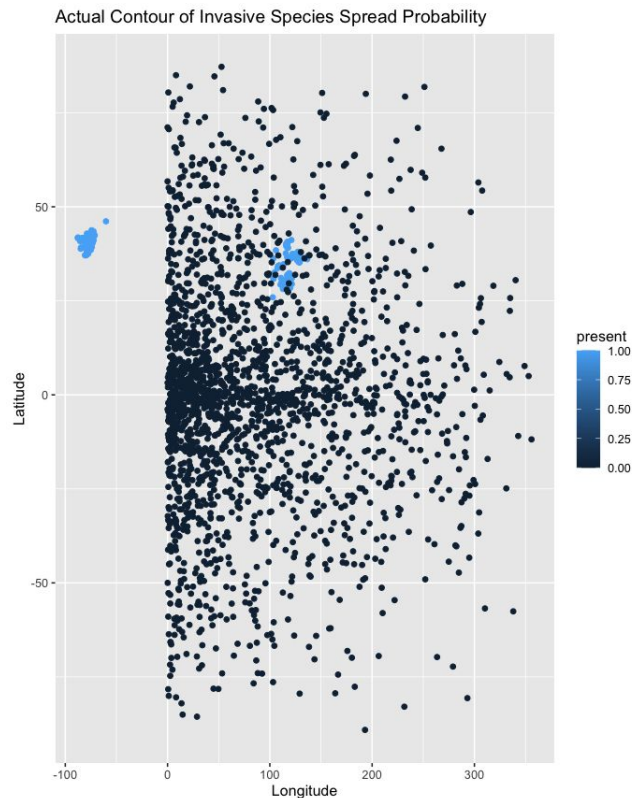
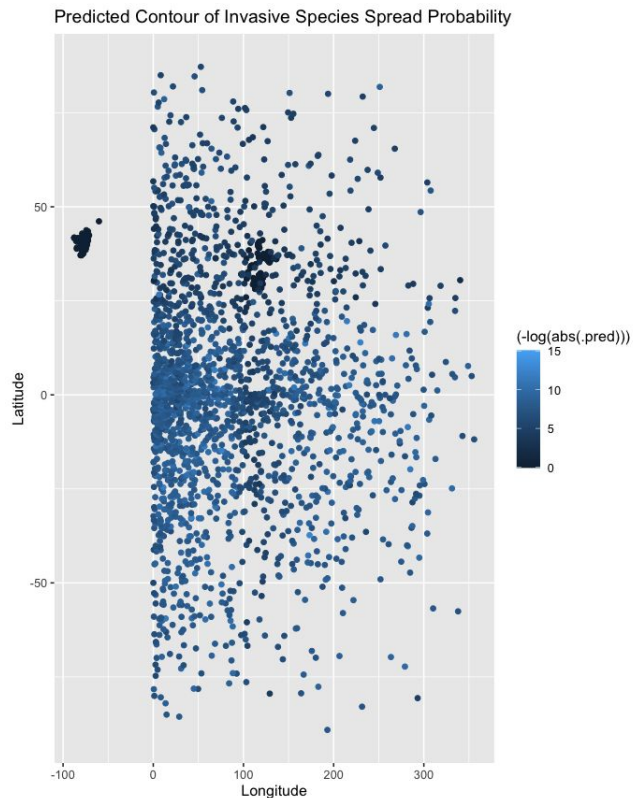


# Our Other Species

*Lycorma delicatula*

RMSE: 0.00925

$R^2$ : 1.0

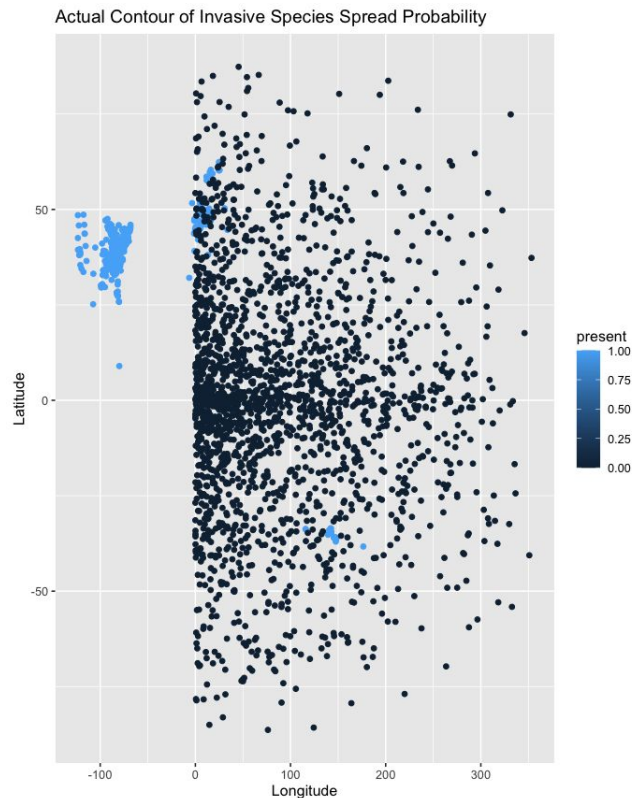
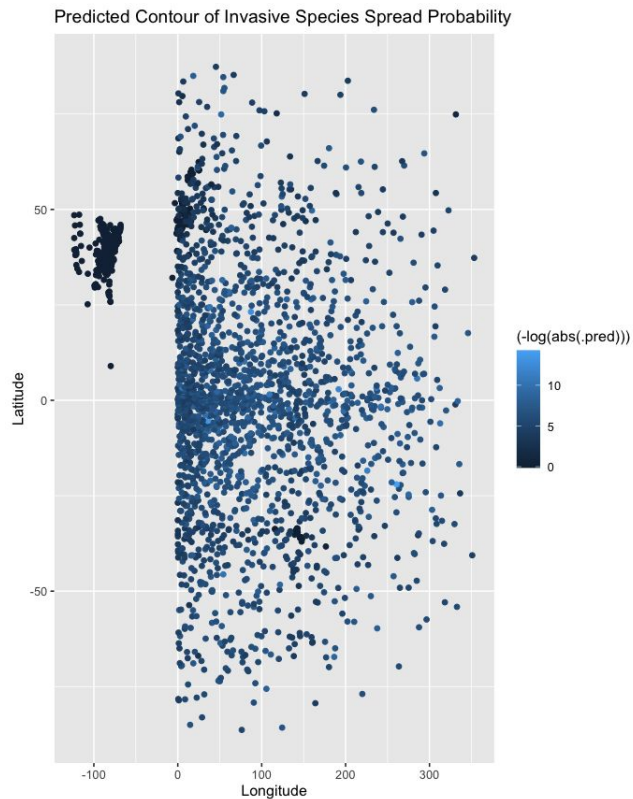


# Our Other Species

*Craspedacusta  
sowerbii*

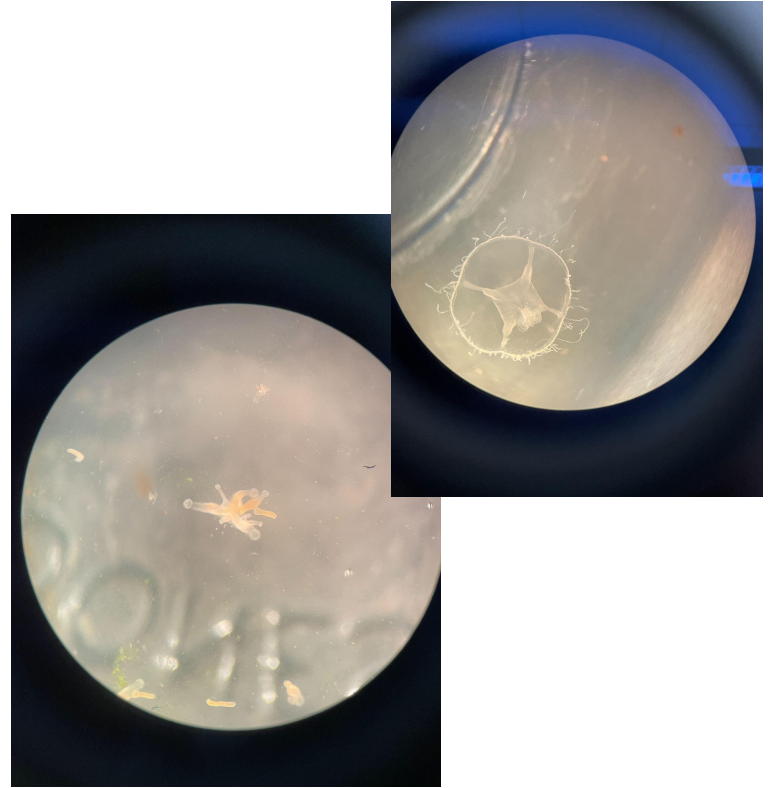
RMSE: 0.0255

$R^2$ : 0.996



# Conclusions + Next Steps

- My next steps:
  - Compare to the biomod2 projections
  - Make “better” pseudoabsences; eliminate overlap
- Other next steps (Longer-Term):
  - Include more predictor variables
  - Validate models with field data
  - Connect model results to potential biological explanations



*Craspedacusta sowerbii* polyps and medusa grown in a laboratory setting.  
Images courtesy of Jonathan Zhu

# References

Mainali, K. P., Warren, D. L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G., ... & Parmesan, C. (2015). Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global change biology*, 21(12), 4464-4480.

GBIF.org (13 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.3f7tcw>

GBIF.org (29 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.q3tqpy>

GBIF.org (29 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.wavs55>

GBIF.org (29 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.nbrags>

GBIF.org (29 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.56gqnm>

GBIF.org (29 October 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.d85sdb>

Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.

Performance and Analysis of the constructed analogue method applied to US soil moisture applied over 1981-2001. Huug van den Dool, Jin Huang and Yun Fan, *J. of Geophysical Research*, vol. 108, 2003, p 1-16

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., and Townshend, J.R.G., 2013, High-Resolution Global Maps of 21st-Century Forest Cover Change: *Science*, v. 342, no. 6160, p. 850-853.