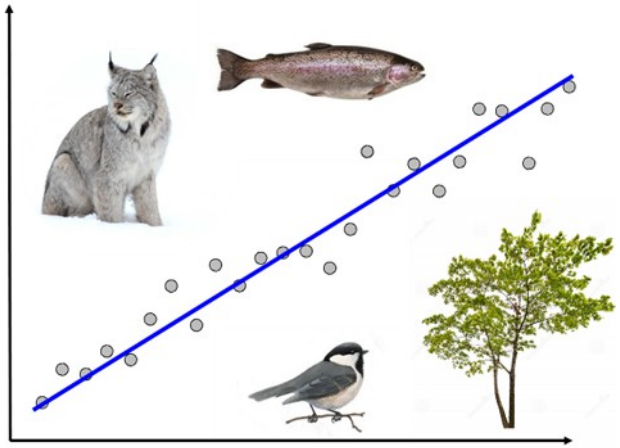
NRC 290b Introduction to Quantitative Ecology

Week 11 – Review and Final Project Introduction



Meg Graham MacLean, PhD

Department of Environmental Conservation

mgmaclean@umass.edu

2019 - Fall

This week

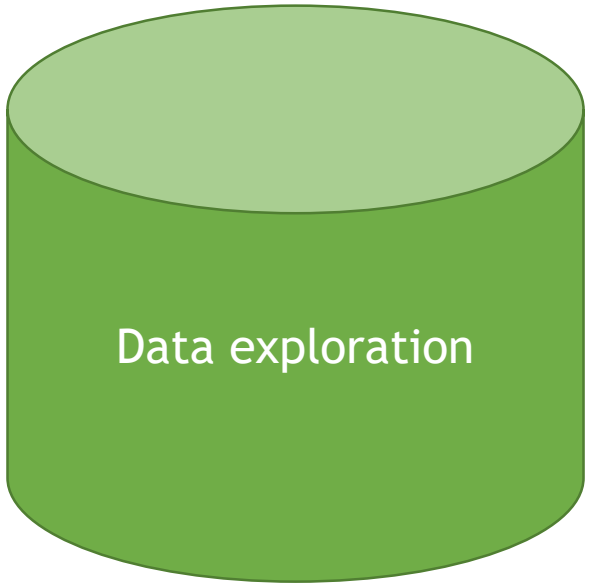
Monday

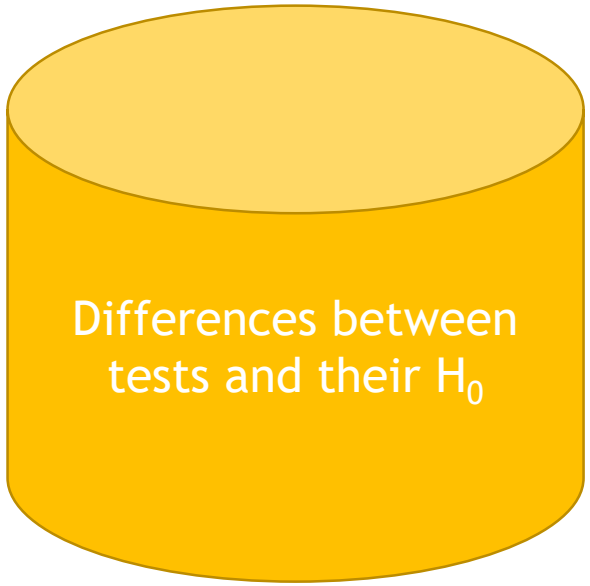
- No class!

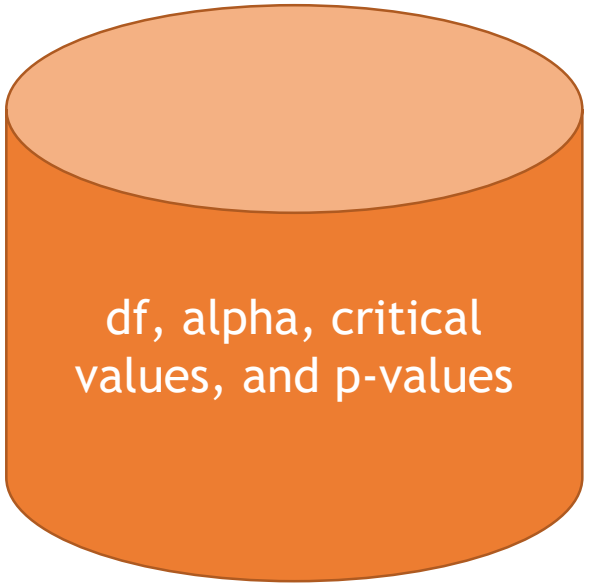
Wednesday

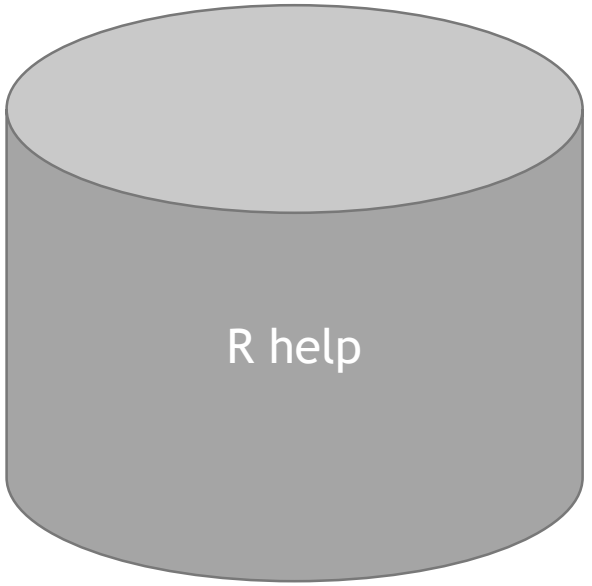
- Review
  - Answer your questions!
- Final Project Introduction
  - Overview
  - Data introduction

Really good questions!









Data exploration

What are the steps to data exploration?

*****THINGS YOU SHOULD DO ANYTIME YOU HAVE NEW DATA*****

1. How were the data collected? (if at all possible to figure out)
    1. Random, systematic, mixed, or haphazard sampling
2. What type of data do you have?
    1. Interval (i.e. continuous), ordinal, or categorical
3. Are your data normally distributed?
    1. Do a Shapiro-Wilk test for normality (H_0: the distribution of the data is not significantly different from normal)
4. What does the distribution of the data look like (using the appropriate methods!)
    1. E.g. range, median, quartiles
    2. For normally distributed data: mean and standard deviation

Differences between tests and H_0

Response variable type	Explanatory variable type(s)	Question type	Test	H_0	Test statistic
Interval	Categorical (2)	Is there a difference between the two response groups?	T-test or U-test*	There is no sig. difference between the means/medians of the two groups	t
Interval	Categorical (2+), can have >1	Is there a difference between the two response groups?	ANOVA or Kruskal-Wallis*	There is no sig. difference between the means/medians of the two groups	F (And Tukey HSD)
Categorical	Categorical	Are the two variables linked? (or is there a dif between obs and exp?)	Chi-square	There is no sig. association between variables (or no dif. between observed and expected)	x^2
Interval	Interval (1 variable)	Are the two variables linked?	Pearson or Spearman*	There is no sig. correlation between variables	r
Any	Any and any number of variables	Do the explanatory variables explain the response variable?	Regression	The slope of the linear relationship is no different from 0	R^2

* For skewed data

df, alpha, critical values, and p-values

- Degrees of freedom (df) is the number of samples you are using to conduct the test – 1
  - -1 is a small sample size penalty
  - Differs depending on the test you are conducting
- P-value is the probability of finding your measured result if H_0 is true
- Alpha is your chosen significance value, or how much "risk" you are willing to accept (usually 0.05)
- Critical values are specific to the test statistic, but are the calculated test statistic value for a given alpha level
  - You can find all of these tables in the back of your book starting on pg. 387 (Appendix 2)

df, alpha, critical values, and p-values

R code to demonstrate with t-test

Picking the best regression model

1. Graph each explanatory variable against the response variable
    1. What do the relationships look like? Linear? Something else? Positive/negative?
2. Create a model with all of the explanatory variables (and interactions) that **make sense for your study system**
    1. Which are significant? <- you don't always have to drop those that aren't significant, especially if the p-value is close to your alpha level
3. Create a model with the significant explanatory variables
    1. Or any you think remain important
4. Use AIC and forward/backward stepwise regression (add1(), step(), or manually) to find the model with the lowest AIC

Picking the best regression model

```
bunnies.glm <- glm(formula = Floppy ~ Weight + BodyLength, family = "binomial", data =
bunnies)
step(bunnies.glm)
Start:  AIC=843.1
Floppy ~ Weight + BodyLength

             Df Deviance    AIC
<none>            837.10 843.10
- BodyLength  1   839.95 843.95
- Weight      1   846.25 850.25

Call: glm(formula = Floppy ~ Weight + BodyLength, family = "binomial", data = bunnies)

Coefficients:
(Intercept)       Weight   BodyLength
      1.730       -3.735        1.037

Degrees of Freedom: 613 Total (i.e. Null);  611 Residual
Null Deviance:     849.9
Residual Deviance: 837.1      AIC: 843.1
```

R Help

- I really like cheat sheets for quick reminders! (see R help section on moodle for some examples)
- R help files
- Other sites:
  - Stack overflow
  - Sthda.com
  - Data camp
  - R programming wikibook?
  - Lynda?
  - R-pkgs.org (for the ambitious R user!)
  - R for data science (free book), Hadley & Wickem
  - The art of R programming (book)
- Anything else you have encountered?

Statistical testing

When doing a regression – what is the H_0 the p-value is testing?

a) The slope is no different from 0

b) The explanatory variable is no different than the response variable

c) The explanatory variable is significantly correlated with the response variable

d) There is no significant difference between groups

Final Project

Let's practice with some real data!

Two parts to the final project:

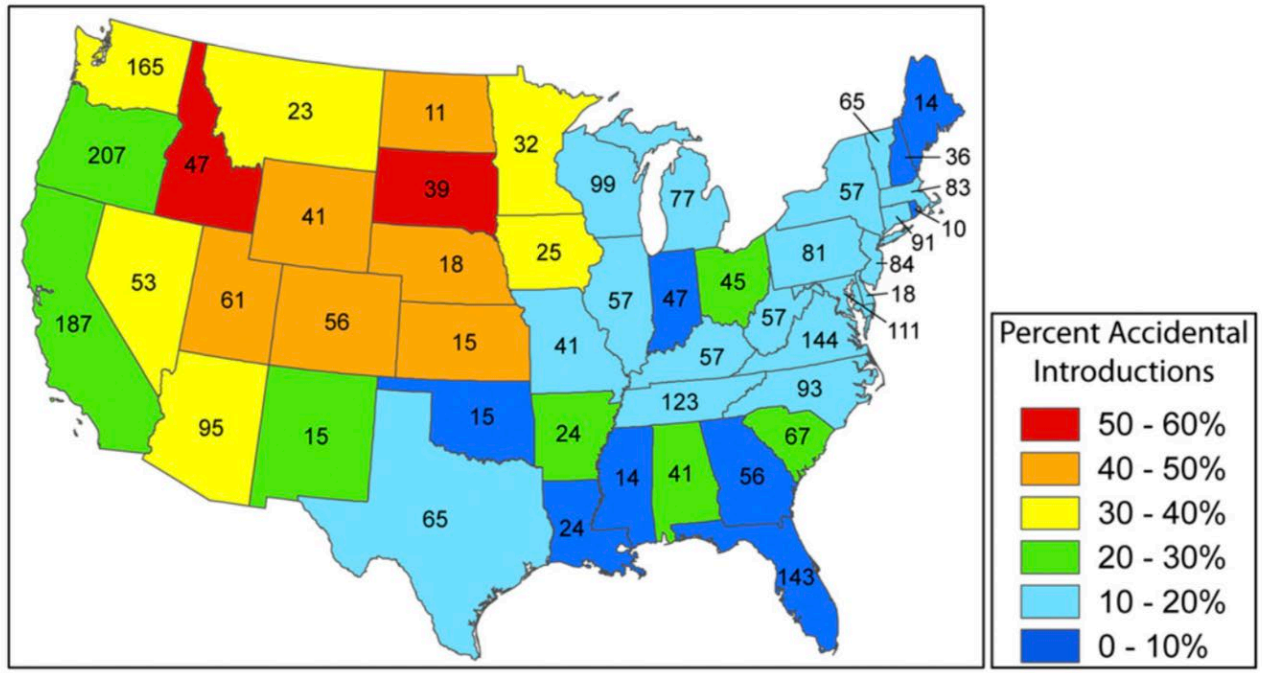
- **Presentations (30% of grade)**
  - 7 minutes to present preliminary data/results, 3 for questions/suggestions for improvement
  - Wednesday, December 11th, in class, **slides due the night before!**
- **Report (70% of grade)**
  - Structured after a peer-reviewed scientific journal article
  - Plus .R script
  - Due by 11:55pm on Wednesday, December 18th

Dataset 1 – Invasive plant introductions

ACCIDENTAL INTRODUCTIONS ARE AN IMPORTANT SOURCE OF INVASIVE PLANTS IN THE CONTINENTAL UNITED STATES[1]

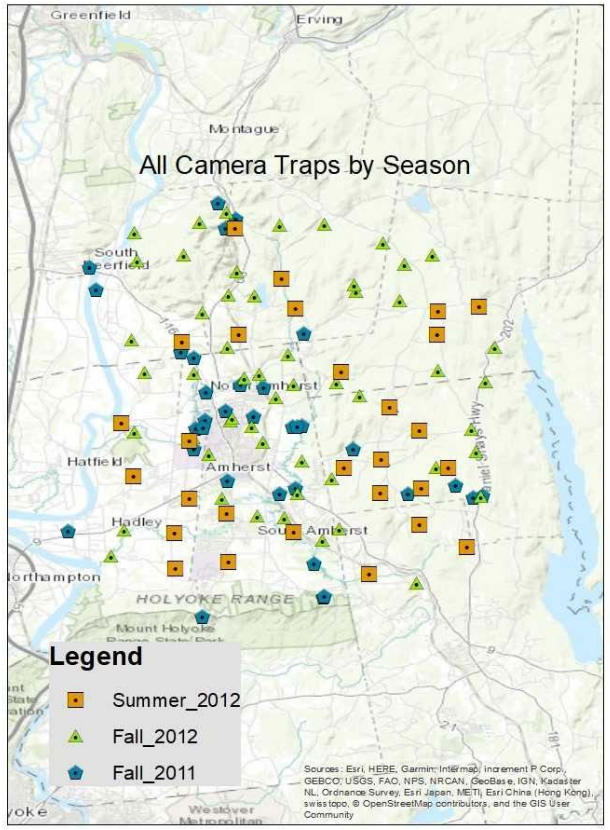
NORA E. LEHAN[2], JULIA R. MURPHY[2], LUKAS P. THORBURN[3], AND BETHANY A. BRADLEY[2,4]

[2]Department of Environmental Conservation, University of Massachusetts, Amherst, Massachusetts 01003 USA; and
[3]Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003 USA



Dataset 2 – Camera traps

Mammal species presence data from camera traps around Amherst, MA



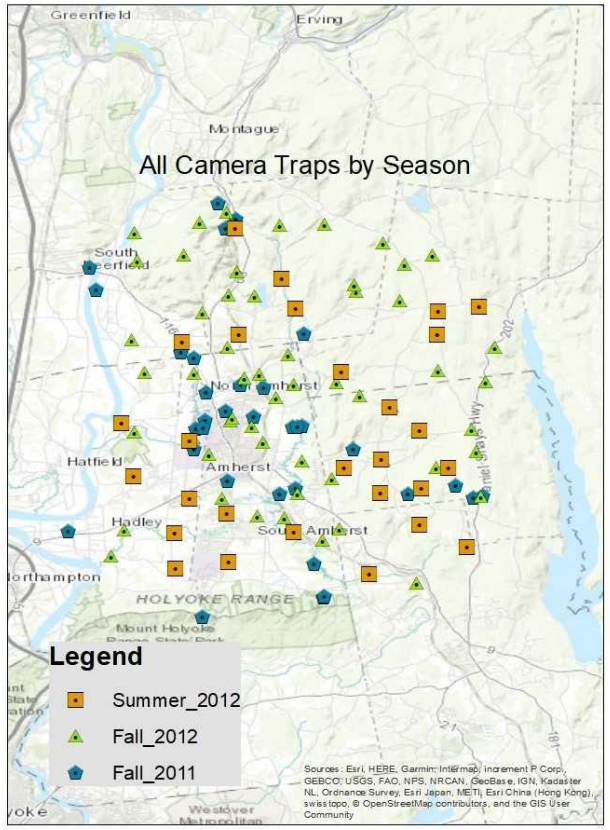
2019 *NORTHEASTERN NATURALIST* 26(2):325–342

Wild Canid Distribution and Co-existence in a Natural–Urban Matrix of the Pioneer Valley of Western Massachusetts

Eric G. LeFlore[1], Todd K. Fuller[1,*], John T. Finn[1], Stephen DeStefano[2], and John F. Organ[3]

Dataset 3 – Forest for the birds

Forest and bird data for many of the same locations as the mammal data (could combine!)



Edited for NRC 564, Fall 2019

Bird Habitat Inventory Field Procedures

This protocol was designed by Audubon Vermont biologists and Vermont foresters to be used to monitor changes in forest bird habitat and timber resource in forest stands over time. However, it may also be used as an adaptable template for foresters who want to integrate inventory of bird habitat features into a traditional forest inventory or timber cruise.

Dataset 4 – Snowstorm tree failure

RESEARCH Open Access

Factors affecting branch failures in open-grown trees during a snowstorm in Massachusetts, USA

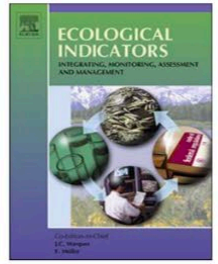
Brian Kane* and John T Finn

Dataset 5 – Invasive bee sampling

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

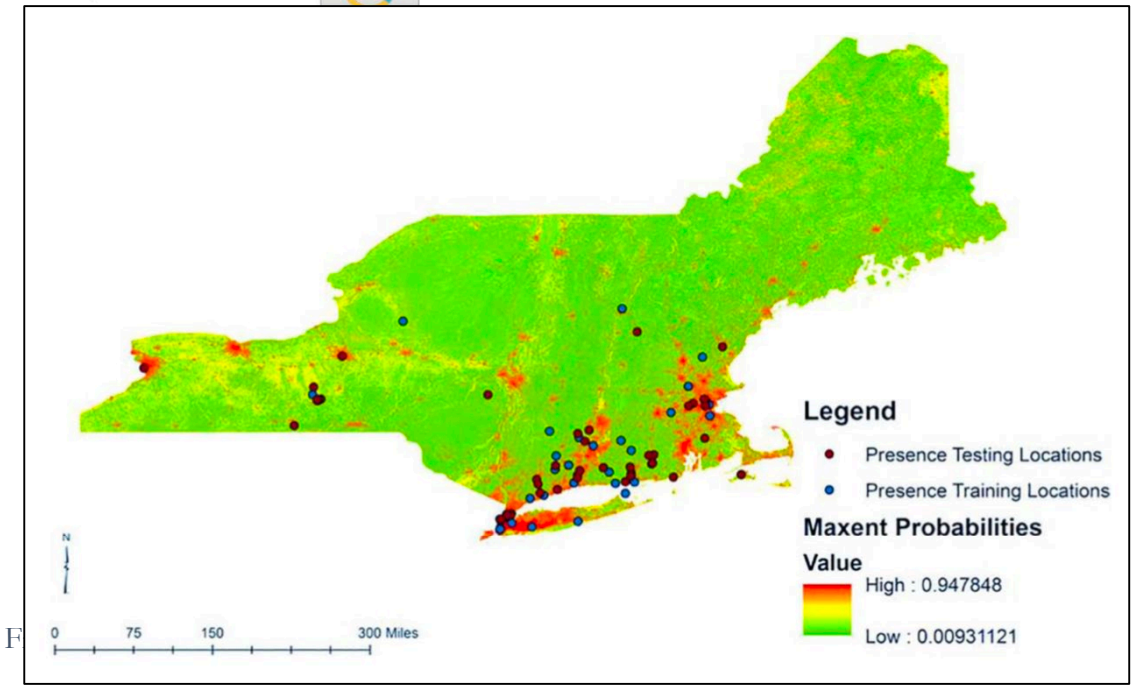


Original Articles

Presence-only modeling is ill-suited for a recent generalist invader, *Anthidium manicatum*

Kelsey K. Graham[a,*], Meghan Graham MacLean[b]

[a] Tufts University, Biology Department, 163 Packard Avenue, Medford, MA 02155, United States
[b] Babson College, Division of Math and Science, 231 Forest Street, Babson Park, MA 02457-0310, United States



Dataset 6 – Oysters

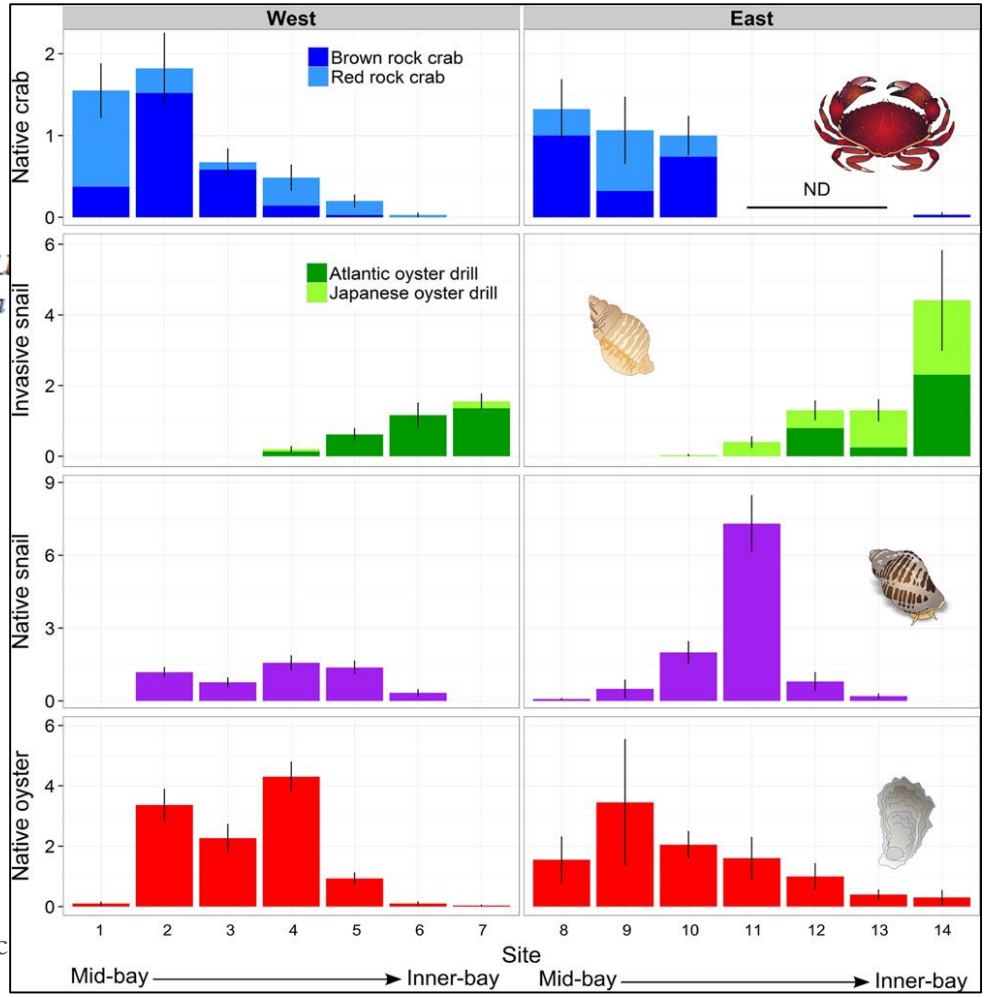
esa

ECOSPHERE

Environmental stress mediates trophic cascade strength and resistance to invasion

BRIAN S. CHENG[1,2,3,†] AND EDWIN D. GROSHOLZ[1,2]

[1]*Bodega Marine Laboratory, University of California, Davis, 2099 Westside Road, Bodega Bay, California 94923 U*
[2]*Department of Environmental Science and Policy, University of California, Davis, One Shields Avenue, Davis, California*



Dataset 7 – Forest composition

FIA forest data from Massachusetts (there are lots of options here!)

For Monday:

1) Read chapter 13 in your book
2) Read Houle et al. (2015) Impacts of Climate Change on the Timing of the Production Season of Maple Syrup in Eastern Canada
3) Pick out your own peer-reviewed journal article to assess.
4) Complete the reading assessment on moodle

All before 11:55pm on Sunday