Independent Analysis Project

- **The project consists of six parts:**

Project Proposal	5
Data Screening	5
Project Presentation	5
Project Analysis	10
Project Report	10
Project Logbook	5
Peer Review	10
Project Total	**40**

Independent Project

- **The project consists of five parts:**

Project Proposal	
Data Screening	April 22
Project Report & Logbook	May 1
Project Analysis & Presentation	May 6
Project Peer Review (Answers)	May 8

Project Assignment – Approach

Follow three steps:

1. Explore species abundance / environmental data

   Check cross-correlations of environmental data
   Check data distributions (skewness, zeros) and outliers
   Implement data transformations / relativization – if needed

2. Analyze data using MVS and report results
(as if for a publication) using text, tables, plots

3. Complete peer review:
   your own work – reanalyze
   another student's work – read and grade

Project Presentation Template (May 6)

- Make a 12 minute presentation of your results
  (~ 82 mins for the entire class)
  (7 students on each day)
  NOTE: send ppt by 9 pm on May 5 to:
  khyrenba@gmail.edu

- **DO NOT HAVE MORE THAN 12 SLIDES**

- **NOTE:** You will get questions from other students and will provide answers in your peer review

Slide 1: Objective

- Explanatory Title – Include brief description of objectives and specific approach (methods used)

- Provide “Goal Statement” (from proposal):

  “The goal of this project was to …”

- Explain relevant hypotheses / predictions

Slide 2: Dataset Description

- Data file: PCA1M.wk1 (main matrix)

```
      96 samples
       5 variables
         Q          Q          Q          Q          Q
         time       MEI        PDO        up_36      up_39
1997jan      1997     -0.47      0.23        -21      -0.27
```

- 96 samples and 5 variables
- Samples are monthly values (Jan. 97 - Dec. 04)
- Variables:

  - Time: decimal year
  - MEI: El Niño Multivariate Index (positive: warm, negative: cold)
  - PDO: Pacific Decadal Oscillation (positive: warm, negative: cold)
  - Up36: upwelling at 36 N (positive: upwelling, negative: downwelling)
  - Up39: upwelling at 39 N (positive: upwelling, negative: downwelling)

Slide 3: Dataset Processing

- How many samples discarded:
  - outliers
  - “empty” samples

- How many species discarded? Describe criteria
- How many samples discarded? Describe criteria

- Outliers: How many columns / rows ?

- If you run into problems with data analysis:
describe data transformations / relativizations used

- Describe your sample size
  samples, species, environmental variables

Slide 4: Dataset Exploration

- Use scatterplot matrix to show plots of pair-wise combinations of the environmental variables

– after data cleaning and transformations

(NOTE: focus on “significant” patterns)

- If there are any cross-correlated environmental variables (visually obvious), describe the sign of the correlation and provide a larger scatterplot
- If you have “time” (year / month / season) as a variable, check for temporal trends (correlations with variables)

Slide 5: Dataset Analysis

- Show the settings used in the analysis by pasting the beginning of the Results.txt file

```
NMS Results
Ordination of stands   in species  space.         20 stands         25 species

         The following options were selected:
ANALYSIS OPTIONS
         1. REL.SOREN. = Distance measure
         2.          6 = Number of axes (max. = 6)
         3.        250 = Maximum number of iterations
         4.     RANDOM = Starting coordinates (random or from file)
         5.          1 = Reduction in dimensionality at each cycle
         6.       0.20 = Step length (rate of movement toward minimum stress)
         7.   USE TIME = Random number seeds (use time vs. user-supplied)
         8.         10 = Number of runs with real data
         9.         20 = Number of runs with randomized data
        10.         NO = Autopilot
        11.   0.000500 = Stability criterion, standard deviations in stress
                         over last 200 iterations.
OUTPUT OPTIONS
        13.         NO = Write distance matrix?
        14.         NO = Write starting coordinates?
        15.         NO = List stress, etc. for each iteration?
        18.        YES = Plot stress vs. iteration?
        17.        YES = Plot distance vs. dissimilarity?
        16.        YES = Write final configuration?
        19.  UNROTATED = Write varimax-rotated or unrotated scores for graph?
        20.        YES = Write run log?
        21.        YES = Write weighted-average scores for species ?
-------------------------------------------------------------------------------
```

Recommendation: Provide the results needed to interpret the results (statistic), the % variance, the axes (descriptions / loadings) and a graphical representation

Slide 6: Results Interpretation

- Examine the criteria for selection of number of axes

Criterion 1: Decline in Stress with added axis at least 5

Criterion 2: P value < 0.05

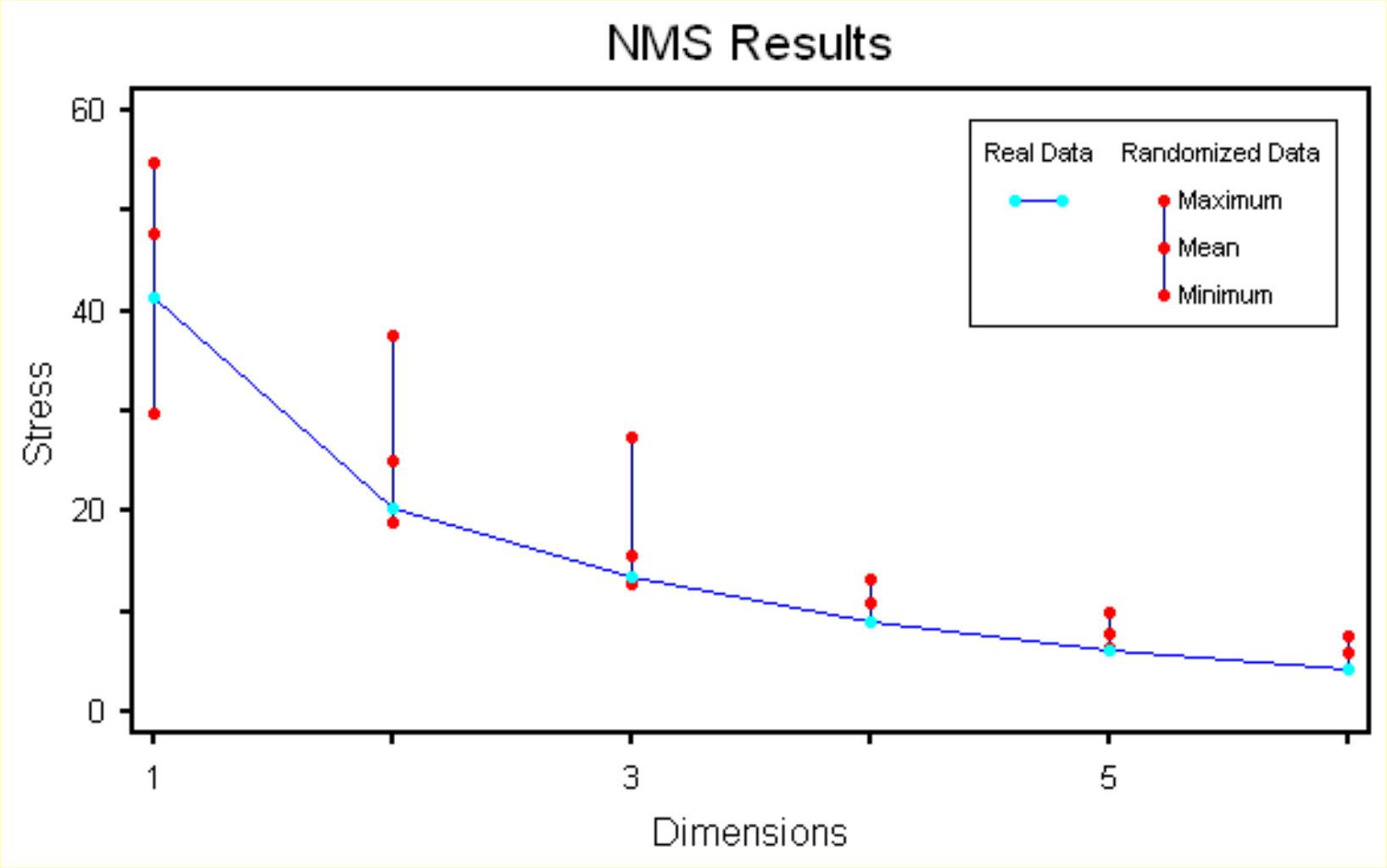
STRESS IN RELATION TO DIMENSIONALITY (Number of Axes)

	Stress in real data 10 run(s)			Stress in randomized data Monte Carlo test, 20 runs			
Axes	Minimum	Mean	Maximum	Minimum	Mean	Maximum	p
1	38.376	46.541	54.222	41.561	48.626	54.483	0.0476
2	20.366	22.469	25.766	21.752	24.574	28.997	0.0476
3	13.418	13.670	14.855	13.809	15.954	17.877	0.0476
4	8.919	8.954	9.268	8.579	10.807	12.085	0.0952
5	6.078	6.288	6.587	6.662	7.863	9.987	0.0476
6	4.138	4.217	4.499	4.635	5.716	7.708	0.0476

p = proportion of randomized runs with stress < or = observed stress
i.e., p = (1 + no. permutations <= observed)/(1 + no. permutations)

Slide 7: Results Interpretation

- Examine and explain the Scree Plot



Slide 8: Results Interpretation

- Coefficient of Determination (% of Variance):

For each axis – together

Coefficients of determination for the correlations between ordination distances and distances in the original n-dimensional space:

Axis	R Squared Increment	Cumulative
1	.126	.126
2	.281	.407
3	.319	.725

- Report Orthogonality:

Measure of independence of the three axes

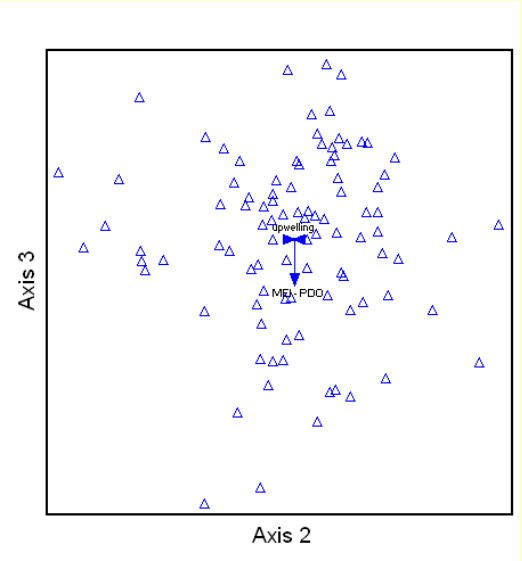
Increment and cumulative R-squared were adjusted for any lack of orthogonality of axes.

Axis pair	r	Orthogonality,% = 100(1-r^2)
1 vs 2	0.021	100.0
1 vs 3	-0.149	97.8
2 vs 3	0.153	97.7

Slide 9: Results Interpretation

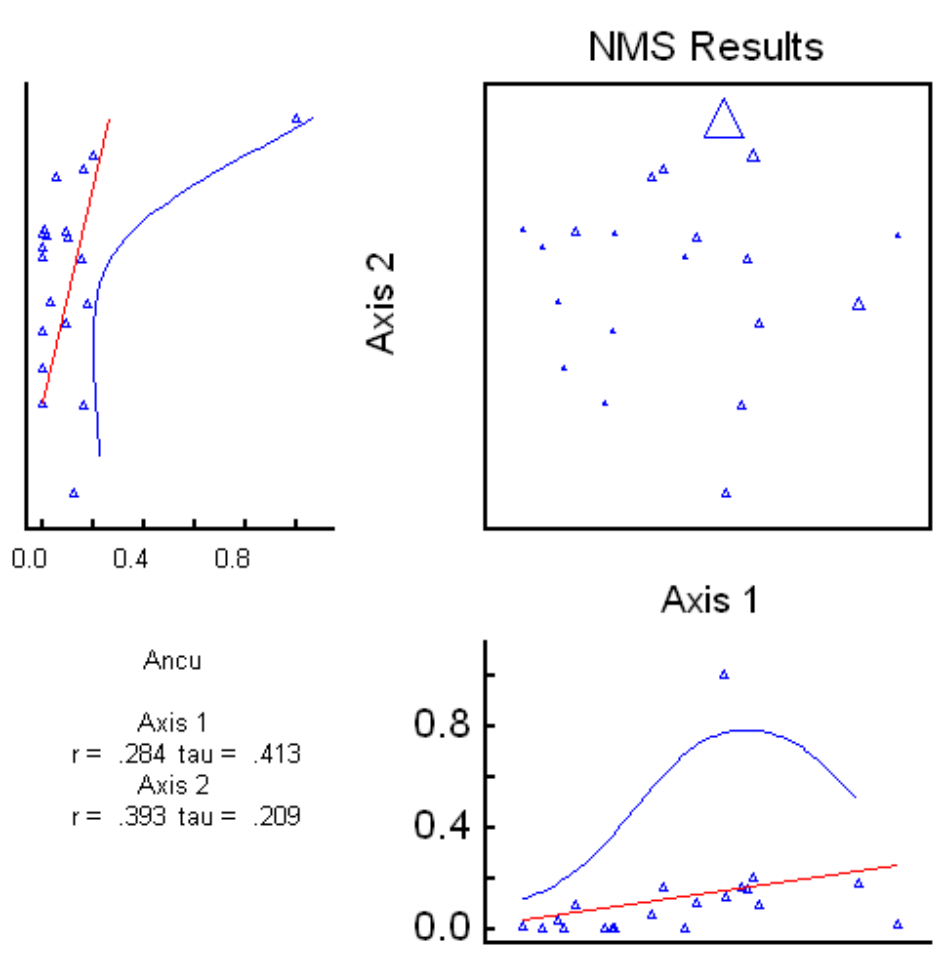
- Examine and explain the Ordination Plot
- Explain which plots are you showing and why?
- Explain what axes mean? (use correlations of variables)
- Highlight any outlier samples / species in the plot
- If using environmental variable vectors – explain them

Mention what is the final stress and what does it mean (according to Clarke's 1993 rule of thumb)



Slide 10: Results Interpretation

- Highlight some of the environmental variable correlations with individual NMS axes. If possible, select cross-correlated variables revealed in the data exploration



Slide 11: Discussion – the Method

- What do these results mean for the hypotheses / predictions you proposed ?

- What did this exercise teach you regarding your overarching analysis methods and objective ?

Slide 12: Discussion – Next Steps

- What do you propose to do for your re-analysis ?

- What would be the next steps for this study ?

Project Assignment – Report (May 1)
No Minimum Length Limit – 3 Sections

- **Methods: (5 points)** Describe with words, plots & tables
  - How many species selected for analysis ? Criteria used?
  - How many environmental variables used? Are they Q / C?
  - Discuss correlations of environmental variables
  - Summarize data manipulations / transformations

- **Results: (5 points)** Describe with words, plots & tables
  - Selection of Axes (2 criteria)
  - Axes correlations with variables
  - Ordination Plots and P values
  - Final variance explained and results

- **Documentation: (5 points)**

- Provide **Log** and 3 files: raw_data, clean_data, results.txt

Log: 3 point, raw/clean data: 1 point, results: 1 point

Project Assignment – Report

Add four or more slides at the end of your ppt

- **1. Methods:** Describe in words – cite plots / tables
  - Discuss correlations of environmental variables

- **2. Results:** Describe in words – cite plots / tables
  - Describe selection of Axes (multiple criteria)

- **3. Results:** Describe in words – cite plots / tables
  - Describe axes correlations with variables

- **4. Results:** Describe in words – cite plots / tables
  - Describe ordination plots

Project Assignment – Report

- **1. Methods:** Describe in words – cite plots / tables
  - Discuss correlations of environmental variables

Project Assignment – Report

- **2. Results:** Describe in words – cite plots / tables
  - Describe selection of Axes (multiple criteria)

Project Assignment – Report

- **3. Results:** Describe in words – cite plots / tables
  - Describe axes correlations with variables

Project Assignment – Report

- **4. Results:** Describe in words – cite plots / tables
  - Describe ordination plots

Project Assignment – Peer Review (May 8)

Reviewer Name: _____

- **Questions / Answers:**

- Copy Questions – Provide 5 Answers **(0.5 point each)**

- **Review: (0.5 point each)**

- Read your partner's ppt and address the following:

- Selection of number / identify of analyzed species
- Data transformations / manipulations
- Selection of the number of axes
- Interpretation of axes (env. correlations)
- Interpretation of variance explained

NOTE: If you cannot answer because there is not enough data… explicitly state so in your review

Project Assignment – Peer Review

- **Re-analysis:** Do another analyses of your own data after consulting with me

- Report the standard outputs (listed in class) for the analysis you completed. (2.5 points)

- Write a paragraph comparing results from both analyses (NMS vs other). (2.5 points)

Hint: if possible, focus on comparing a hypothesis-driven vs a more exploratory approach

Project Assignment – Peer Review

Reviewer Name: _____

- **Questions / Answers:**

Copy Questions – Provide 5 Answers **(0.5 each)**

(Add more slides – if needed)

Project Assignment – Peer Review

Reviewer Name: _____

- **Review: (0.5 points each)**

Read your partner's ppt and address the following:

- Selection of number / identify of analyzed species
- Data transformations / manipulations
- Selection of the number of NMS axes
- Interpretation of NMS axes (env. correlations)
- Interpretation of variance explained / final stress

(Add more slides – if needed)

Project Assignment – Peer Review

- **Re-analysis:** Do another analyses (ideally, hypothesis driven) after consulting with me

- Report the standard outputs (listed in class) for the analysis you completed (2.5 points)

- Write paragraph comparing results from both analyses (exploration vs hypothesis) (2.5 points)
  Hint: You may want to focus on environmental associations and clusters of stations / species

(Add more slides – if needed)