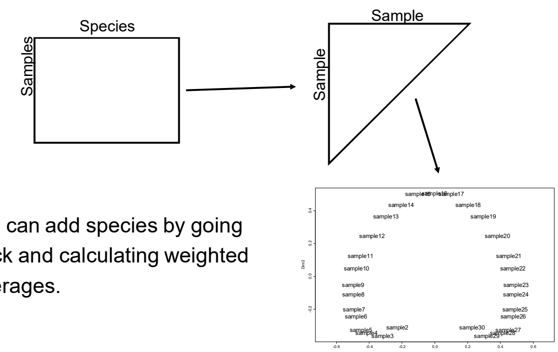
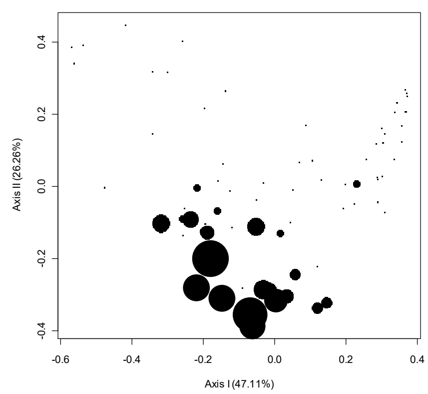
PCoA – Where are the species?

- Because PCoA is based on a distance matrix, the analysis never “sees” any species data.



We can add species by going back and calculating weighted averages.

Notropis texanus overlay



Species Weighted Averages

From the original data matrix (species abundance): for a given species, average the product of that species abundance and axis scores for all sites. Repeat for each species and axis combination.

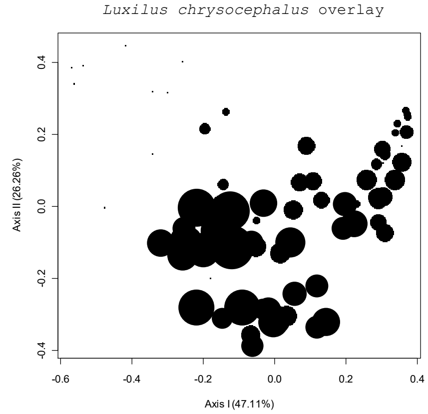
	Axis 1 Scores
Big1_2_2006	-0.146599614
Big2_1_2006	-0.017012001
Big2_1_2006b	0.056843026
Big2_10_2005	-0.062377827
Big2_11_2005	-0.218488715
Big2_2_2006	-0.003900590
Big2_5_2005	0.003548039
Big2_6_2005	0.032613154
Big2_7_2005	-0.033865696
Big2_8_2005	-0.089449436
Big2_9_2005	-0.067947807
Bouie1_3_2006	-0.197247028
Bouie1_7_2005	-0.124918607
Etc…	

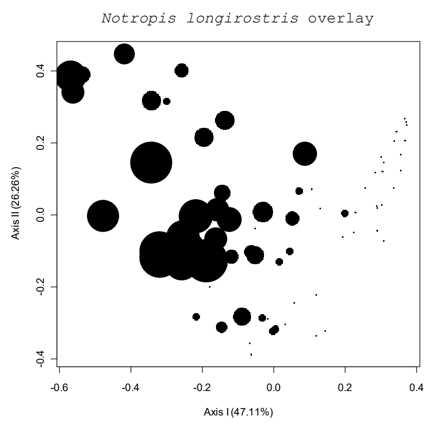
	Notropis.texanus
Big1_2_2006	16
Big2_1_2006	4
Big2_1_2006b	2
Big2_10_2005	15
Big2_11_2005	16
Big2_2_2006	1
Big2_5_2005	11
Big2_6_2005	3
Big2_7_2005	6
Big2_8_2005	0
Big2_9_2005	38
Bouie1_3_2006	0
Bouie1_7_2005	0
Etc…	

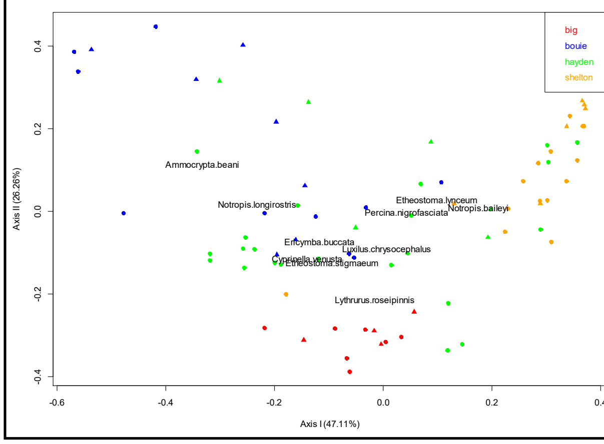
Big1_2_2006	-2.34559383
Big2_1_2006	-0.06804800
Big2_1_2006b	0.11368605
Big2_10_2005	-0.93566740
Big2_11_2005	-3.49581944
Big2_2_2006	-0.00390059
Big2_5_2005	0.03902843
Big2_6_2005	0.09783946
Big2_7_2005	-0.20319418
Big2_8_2005	0.00000000
Big2_9_2005	-2.58201667
Bouie1_3_2006	0.00000000
Bouie1_7_2005	0.00000000
Etc..	

Weighted average for *N. texanus* on axis 1 = -0.0969

Luxilus chrysocephalus overlay







Calculating Weighted Averages

- Function wascores(x,y)
  - x: data matrix
  - y: sample scores ($points part of PCoA object)

  - Returns weighted average scores that can be plotted (example in the PCoA answer posted)

- Note that you could also use wascores() to plot environmental variables…anything for which you have data for each sample.

Principal Components Analysis (PCA)

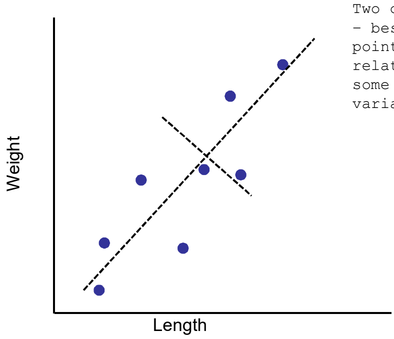
- First and most basic eigenvalue based ordination

- Works with the original dataset, not a distance matrix

- Eigenvalue decomposition of **correlation or covariance** matrix
  - Relationship among variables and gradients is important
  - Relationship to variables is not lost (no need for wascores)

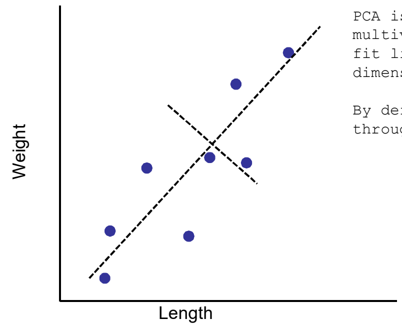
- Works best with linear relationships among variables, thus, some data are not appropriate (species) and others require transformation

Principal Components Analysis (PCA)



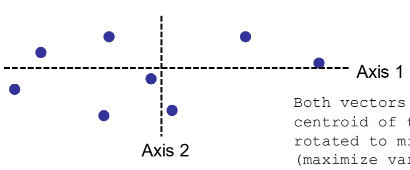

```
Two dimensional regression
– best fit line through
points describes the
relationship, explains
some proportion of
variance.
```

Principal Components Analysis (PCA)




```
PCA is similar but in
multivariate space. Best
fit line in multiple
dimensions = component.

By definition passes
through centroid.
```




```
Both vectors pass through the
centroid of the data and are
rotated to minimize residuals
(maximize variance
explained).

Relative spatial positioning
among points unchanged.

Axis 1 longer = more variance
explained.
```

PCA Functions in R

- Three functions in R:
  - **prcomp()** Singular value decomposition of variance/covariance matrix
    - variance calculated using N-1
    - By default, data are not scaled
  - **princomp()** eigen analysis of wither the correlation of variance/covariance matrix
    - Variance calculated differently (N)
    - No option for scaling data (need to do this before)
    - Option cor specifies if you use correlation or covariance matrix
  - **rda()** vegan package function that will be used for constrained ordinations
    - Similar to prcomp(), species and site scores are rescaled according to Legendre, P. and Legendre, L. (1998) *Numerical Ecology*. 2nd English ed. Elsevier

	sp1	sp2	sp3	sp4	sp5
sample2	0.2588190	0.0000000	0.0000000	0.0000000	0.000000
sample3	0.5000000	0.0000000	0.0000000	0.0000000	0.000000
sample4	0.7071068	0.2588190	0.0000000	0.0000000	0.000000
sample5	0.8660254	0.5000000	0.0000000	0.0000000	0.000000
sample6	0.9659258	0.7071068	0.2588190	0.0000000	0.000000
sample7	1.0000000	0.8660254	0.5000000	0.0000000	0.000000
sample8	0.9659258	0.9659258	0.7071068	0.2588190	0.000000
sample9	0.8660254	1.0000000	0.8660254	0.5000000	0.000000
sample10	0.7071068	0.9659258	0.9659258	0.7071068	0.258819
sample11	0.5000000	0.8660254	1.0000000	0.8660254	0.500000

Raw data

	sp1	sp2	sp3	sp4	sp5	sp6	sp7
sp1	0.143232167	0.116915728	0.058090217	-0.004409783	-0.049052640	-0.068661144	-0.071053547
sp2	0.116915728	0.143232167	0.116915728	0.058090217	-0.004409783	-0.049052640	-0.068661144
sp3	0.058090217	0.116915728	0.143232167	0.116915728	0.058090217	-0.004409783	-0.049052640
sp4	-0.004409783	0.058090217	0.116915728	0.143232167	0.116915728	0.058090217	-0.004409783
sp5	-0.049052640	-0.004409783	0.058090217	0.116915728	0.143232167	0.116915728	0.058090217
sp6	-0.068661144	-0.049052640	-0.004409783	0.058090217	0.116915728	0.143232167	0.116915728
sp7	-0.071053547	-0.068661144	-0.049052640	-0.004409783	0.058090217	0.116915728	0.143232167
sp8	-0.071053547	-0.071053547	-0.068661144	-0.049052640	-0.004409783	0.058090217	0.116915728
sp9	-0.071053547	-0.071053547	-0.071053547	-0.068661144	-0.049052640	-0.004409783	0.058090217
sp10	-0.071053547	-0.071053547	-0.071053547	-0.071053547	-0.068661144	-0.049052640	-0.004409783

Variance-covariance matrix

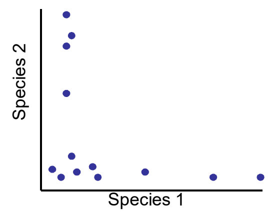
Eigenvalues and eigenvectors. N=smaller dimension of original matrix

Eigenvalues 0.6541252090 0.4627584823 0.14677097090.1363433693 0.0140294277 0.0082396017 0.0040938924

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	-0.3284840	0.24534106	0.4912923	0.17561282	-0.51298964	0.3559863	0.26482374	-0.15467924
[2,]	-0.3952926	0.09482018	0.1444854	0.47885425	-0.14527535	-0.3843255	-0.24593127	0.40872176
[3,]	-0.3834491	-0.13442310	-0.2420519	0.45062547	0.27605119	-0.2122798	-0.09635715	-0.42837624
[4,]	-0.2797332	-0.37155887	-0.3834218	0.17754887	0.04703883	0.3089102	0.53376620	0.06322656
[5,]	-0.1027528	-0.52411004	-0.1793165	-0.07260737	-0.37056685	0.2916838	-0.27421732	0.34859297
[6,]	0.1027528	-0.52411004	0.1793165	-0.07260737	-0.37056685	-0.2916838	-0.27421732	-0.34859297
[7,]	0.2797332	-0.37155887	0.3834218	0.17754887	0.04703883	-0.3089102	0.53376620	-0.06322656
[8,]	0.3834491	-0.13442310	0.2420519	0.45062547	0.27605119	0.2122798	-0.09635715	0.42837624
[9,]	0.3952926	0.09482018	-0.1444854	0.47885425	-0.14527535	0.3843255	-0.24593127	-0.40872176

PCA restrictions

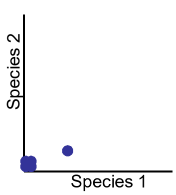
- Relationships among species and gradients are important.
- Thus, variables should be normally distributed and linearly related to gradients.
- Shared zero data is going to be problematic.



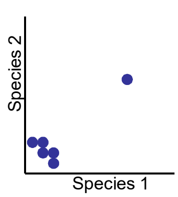
Dust-bunny distributions are common in community data.

No transformation is going to fix that…don't use PCA.

Shared Zeroes and Outliers



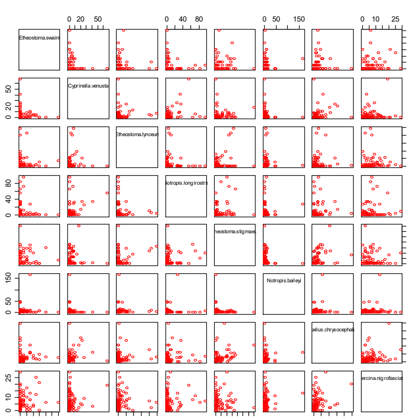
Shared zero data among species will lead to a false high positive relationship.



A single large outlier can define the linear relationship among two variables.

Data Assumptions

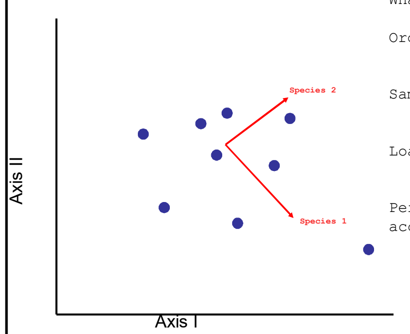
- Often helpful to look at scatter plots of data first
- Common transformations – eliminate rare species, log transform, scale data etc.



Principle Components Analysis (PCA)

- Code
  - `prin_comp<-rda(community, scale=TRUE)`
- Options
  - Scale = rescale data automatically to have unit variance (~ 1SD per variable)
- Output
  - Eigenvalues and eigenvectors
  - Summary of % variance accounted for ("inertia" = % variance)
    - If scale=TRUE then you have 1 unit of inertia per variable
    - If scale=FALSE then variance contributed is not equal across variables
  - Scaling component
  - Species scores
  - Sample scores

Biplot – species and samples in ordination space



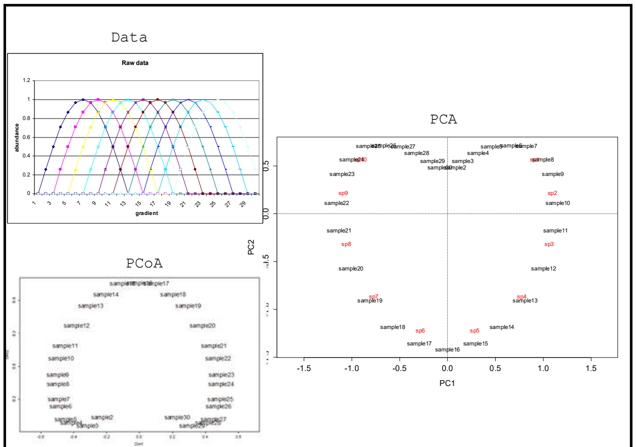
What to present:

Ordination –

Sample Scores –

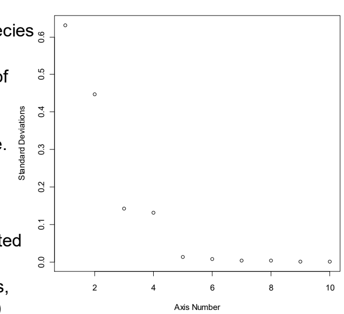
Loadings –

Percent variance accounted for –



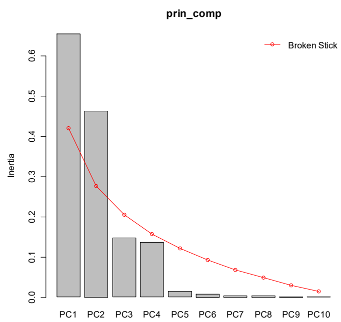
How many axes, which to use?

- By definition
  - # axes = # of variables (species in this case)
  - all axes account for 100% of variance
- First axis accounts for most variance, decreases from there.
- How many to use?
  - Descriptive (2D plot, diminishing returns)
  - *A priori* % variance accounted for
  - Analytical (non-random axis, Monte Carlo determination)

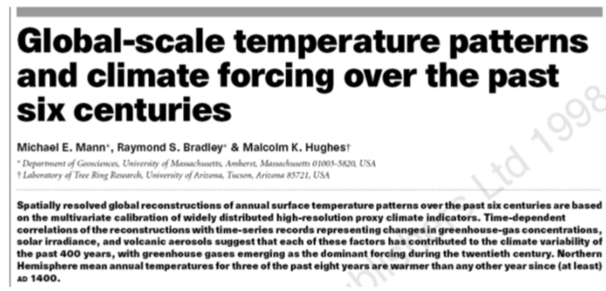


How many axes to use?

- Function screeplot() will plot the variation explained by axis
- Option bstick=TRUE will include the broken stick distribution of variance explained. Broken stick – each axis gets an equal amount of the remaining variation.
- Recommendation is to ignore axes where variance explained is less than broken stick.



Climate change and hockey sticks



Global-scale temperature patterns and climate forcing over the past six centuries

Michael E. Mann*, Raymond S. Bradley* & Malcolm K. Hughes†

* Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003-5820, USA
† Laboratory of Tree Ring Research, University of Arizona, Tucson, Arizona 85721, USA

Spatially resolved global reconstructions of annual surface temperature patterns over the past six centuries are based on the multivariate calibration of widely distributed high-resolution proxy climate indicators. Time-dependent correlations of the reconstructions with time-series records representing changes in greenhouse-gas concentrations, solar irradiance, and volcanic aerosols suggest that each of these factors has contributed to the climate variability of the past 400 years, with greenhouse gases emerging as the dominant forcing during the twentieth century. Northern Hemisphere mean annual temperatures for three of the past eight years are warmer than any other year since (at least) AD 1400.

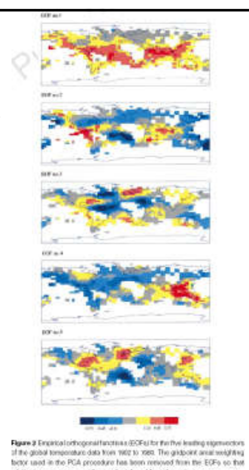
Data used

- Multiple tree ring datasets
- Not directly comparable (differences in local climate, species ecology etc.)
- Data were normalized

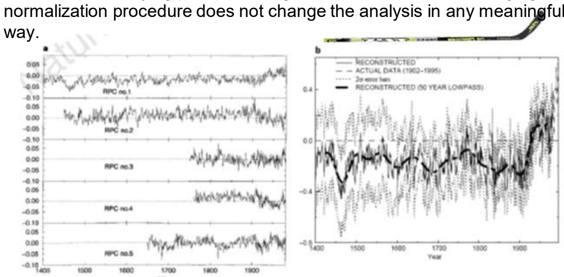
Tree Ring Dataset

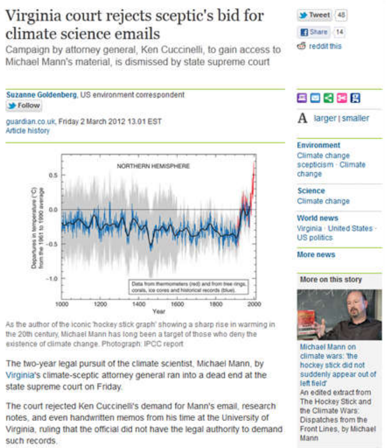
Year

Within each dataset, relative width an indicator of annual mean temperature.



- The most variable tree ring datasets were most closely related to early axes (a function of how PCA works)
- "Criticism" – normalization procedure over emphasized the more variable datasets, the analysis was "fixed" to emphasize recent (anthropogenic) change.
- However – underlying pattern among all datasets is similar, changing the normalization procedure does not change the analysis in any meaningful way.





Virginia court rejects sceptic's bid for climate science emails

Campaign by attorney general, Ken Cuccinelli, to gain access to Michael Mann's material, is dismissed by state supreme court

Tweet 48
Share 14
reddit this

Suzanne Goldenberg, US environment correspondent
Follow

guardian.co.uk, Friday 2 March 2012 13.01 EST
Article history

A larger | smaller

As the author of the iconic 'hockey stick graph' showing a sharp rise in warming in the 20th century, Michael Mann has long been a target of those who deny the existence of climate change. Photograph: IPCC report

The two-year legal pursuit of the climate scientist, Michael Mann, by Virginia's climate-sceptic attorney general ran into a dead end at the state supreme court on Friday.

The court rejected Ken Cuccinelli's demand for Mann's email, research notes, and even handwritten memos from his time at the University of Virginia, ruling that the official did not have the legal authority to demand such records.

Environment
Climate change scepticism · Climate change

Science
Climate change

World news
Virginia · United States · US politics

More news

More on this story

Michael Mann on climate wars: 'the hockey stick did not suddenly appear out of left field'
An edited extract from The Hockey Stick and the Climate Wars: Dispatches from the Front Lines, by Michael Mann

Assignment

- Section 5.3 in the text
- Papers
  - Nash, D., T.D. Als. R. Maile, G. Jones, and J.J. Boomsma. (2008) A mosaic of chemical coevolution in a large butterfly. *Science* **319**: 88-90.
- FYI papers
  - Mann, M.E., R. S. Bradley, and M. K. Hughes. (1998) Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392** (6678):779-787.
  - Peres-Neto, P.R., D. A. Jackson, and K. M. Somers. (2003) Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. *Ecology* **84** (9):2347-2363.
- Example dataset and script

Assignment

- Environmental data describing stream habitat
  - Drainage
  - Variables
    - Cumulative drainage area –
    - pH
    - DO – dissolved oxygen
    - Temp
    - Tds – total dissolved solids
    - Depth
    - Flow
    - Substrate – mean on a 1-6 scale where 1=silt and 6=boulder
    - Cover
    - Vegetation
    - Width
    - Canopy cover
    - Turbidity
- Do a PCA using the rda() function
  - How much variance is explained? How do the variables relate to the axes?
  - Plot the ordination with colors for different drainages.
  - What two variables are best linked to axes 1 and 2?
  - How do the drainages differ?