Ordination

Lecture 09.3: nMDS

Lauren Sullivan

Module: Multivariate Models

Readings

Required for class:

- NA

Optional:

- Strecker, A. L. and Brittain, J. T. (2017) Increased habitat connectivity homogenizes freshwater communities: historical and landscape perspectives. *Journal of Applied Ecology*.
- Dr. Philip Dixon

Multivariate Analysis

There are several ways to look at multivariate patterns from a matrix of **Y**'s.

1. Linear models: MANOVA/regression to test patterns
2. **Ordination: PCA, nMDS, etc to visualize patterns**
3. Permutation tests: PERMANOVA to test patterns

So you have a distance matrix, what's next?

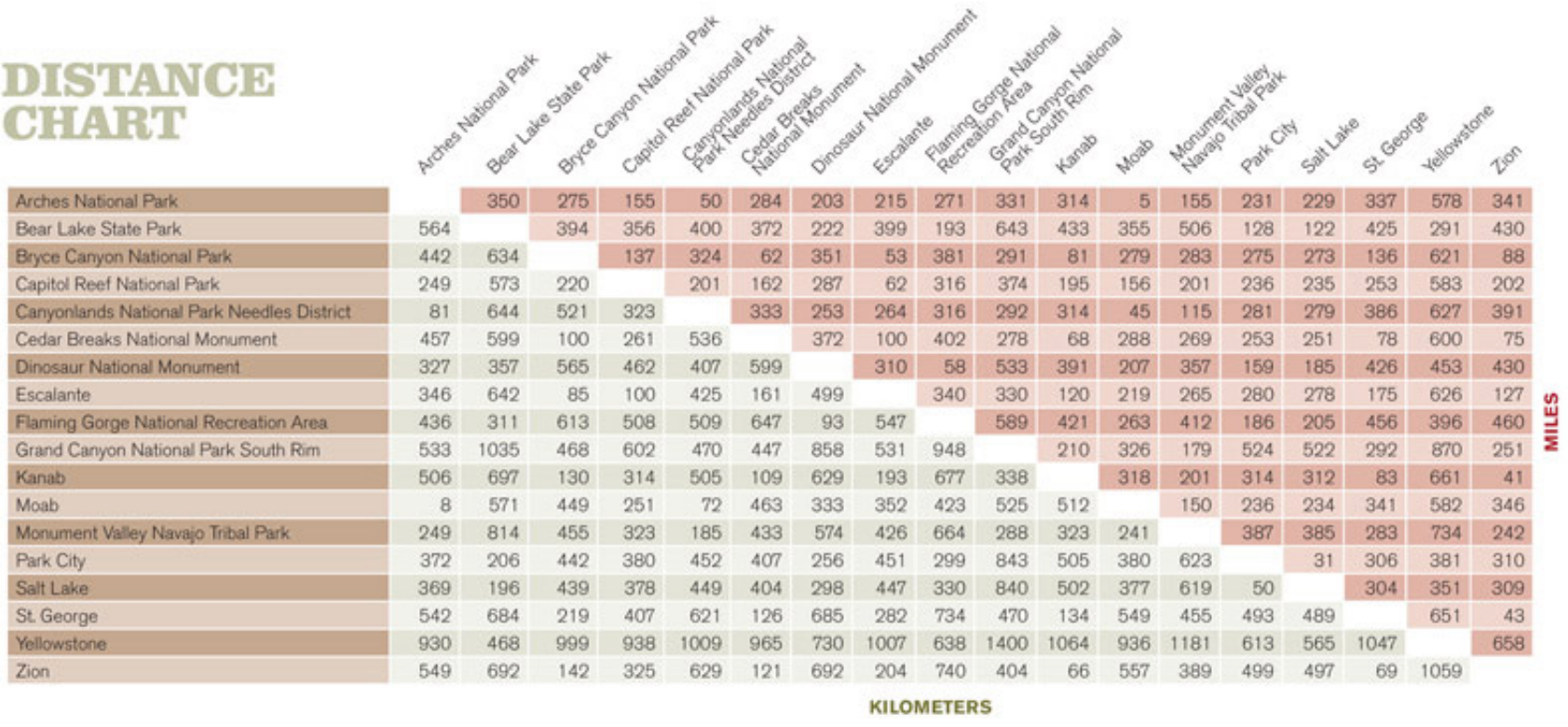
1. Describe variability among replicate sites.
   - Pairwise distances among groups (trts) of sites.
2. Test hypotheses about composition in a site/sample.
   - Compare pairwise distance between reps within and between groups where you'd expect distances to be different.
3. Trends in composition in a site/sample through time.
   - How does composition change through time when you add a treatment?
4. Identify clusters of sites with similar composition ("clustering").
5. Draw pictures that approximate patterns in the distance matrix ("ordination").

nMDS

Non-metric multidimensional scaling represents data in multdimensional space as accurately as possible with a reduced number of dimensions using optimization techniques so patterns can be easily visualized.

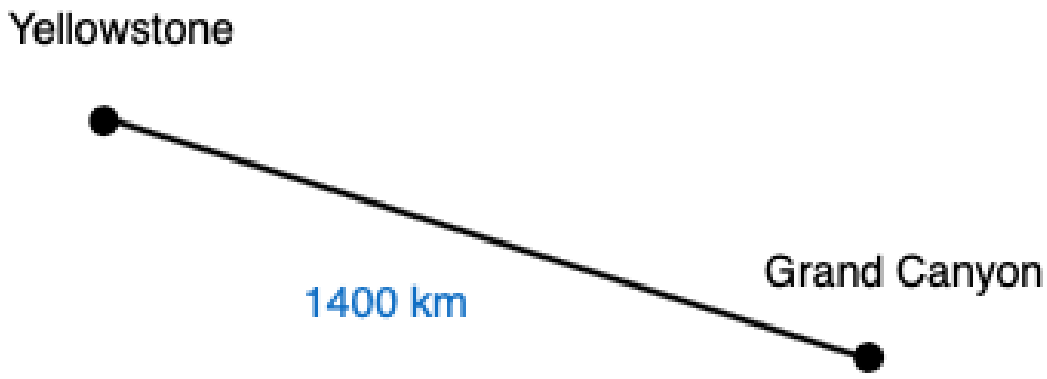
- Unlike PCA (which uses Euclidean distance), nMDS relies on rank orders, or distances, for ordination (this makes it non-metric).
- By using distances to represent differences between sites/samples, you do not have the issues associated with using predictor variables alone (as in PCA).
- This method allows for a variety of data types, and thus is quite flexible!
- Allows you to visualize how environmental gradients influence communities.

How nMDS Works

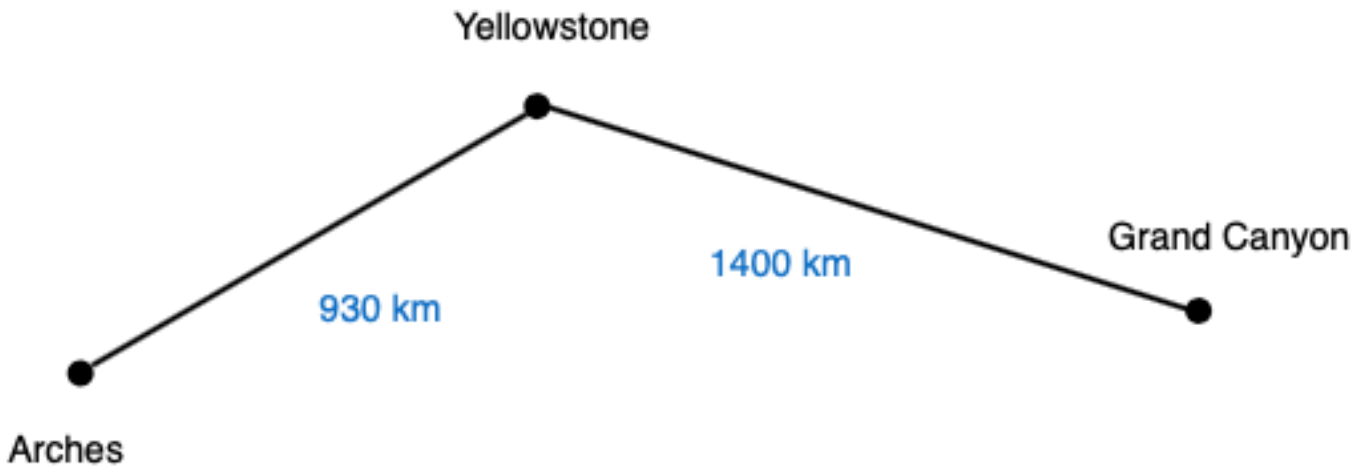


	Arches National Park	Bear Lake State Park	Bryce Canyon National Park	Capitol Reef National Park	Canyonlands National Park Needles District	Cedar Breaks National Monument	Dinosaur National Monument	Escalante	Flaming Gorge National Recreation Area	Grand Canyon National Park South Rim	Kanab	Moab	Monument Valley Navajo Tribal Park	Park City	Salt Lake	St. George	Yellowstone	Zion
Arches National Park		350	275	155	50	284	203	215	271	331	314	5	155	231	229	337	578	341
Bear Lake State Park	564		394	356	400	372	222	399	193	643	433	355	506	128	122	425	291	430
Bryce Canyon National Park	442	634		137	324	62	351	53	381	291	81	279	283	275	273	136	621	88
Capitol Reef National Park	249	573	220		201	162	287	62	316	374	195	156	201	236	235	253	583	202
Canyonlands National Park Needles District	81	644	521	323		333	253	264	316	292	314	45	115	281	279	386	627	391
Cedar Breaks National Monument	457	599	100	261	536		372	100	402	278	68	288	269	253	251	78	600	75
Dinosaur National Monument	327	357	565	462	407	599		310	58	533	391	207	357	159	185	426	453	430
Escalante	346	642	85	100	425	161	499		340	330	120	219	265	280	278	175	626	127
Flaming Gorge National Recreation Area	436	311	613	508	509	647	93	547		589	421	263	412	186	205	456	396	460
Grand Canyon National Park South Rim	533	1035	468	602	470	447	858	531	948		210	326	179	524	522	292	870	251
Kanab	506	697	130	314	505	109	629	193	677	338		318	201	314	312	83	661	41
Moab	8	571	449	251	72	463	333	352	423	525	512		150	236	234	341	582	346
Monument Valley Navajo Tribal Park	249	814	455	323	185	433	574	426	664	288	323	241		387	385	283	734	242
Park City	372	206	442	380	452	407	256	451	299	843	505	380	623		31	306	381	310
Salt Lake	369	196	439	378	449	404	298	447	330	840	502	377	619	50		304	351	309
St. George	542	684	219	407	621	126	685	282	734	470	134	549	455	493	489		651	43
Yellowstone	930	468	999	938	1009	965	730	1007	638	1400	1064	936	1181	613	565	1047		658
Zion	549	692	142	325	629	121	692	204	740	404	66	557	389	499	497	69	1059	

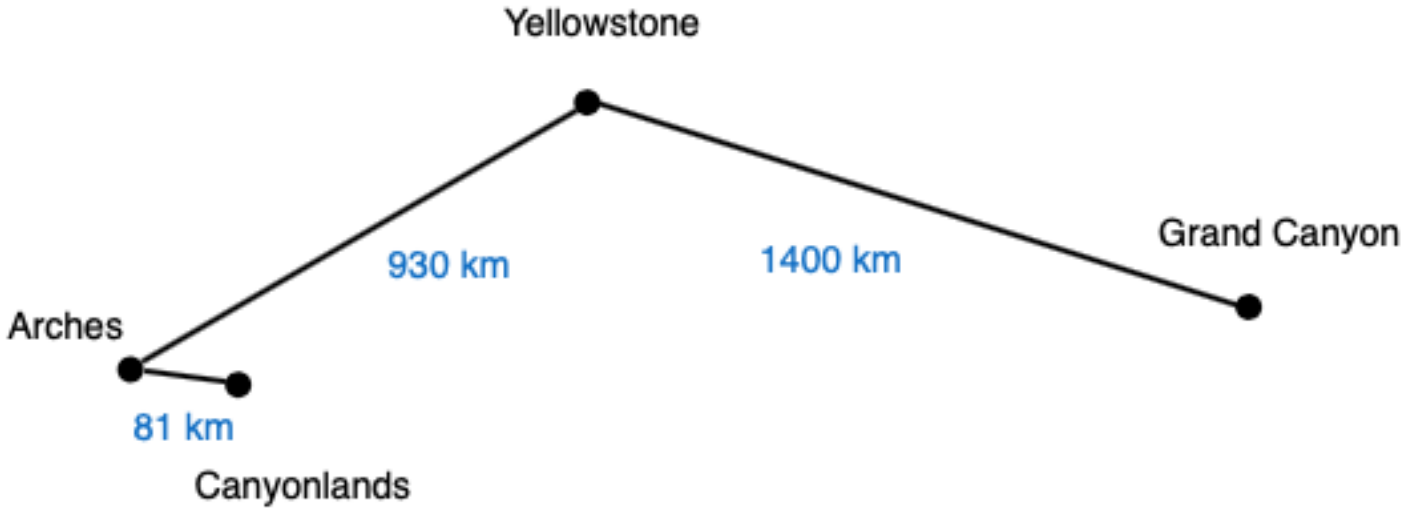
How nMDS Works



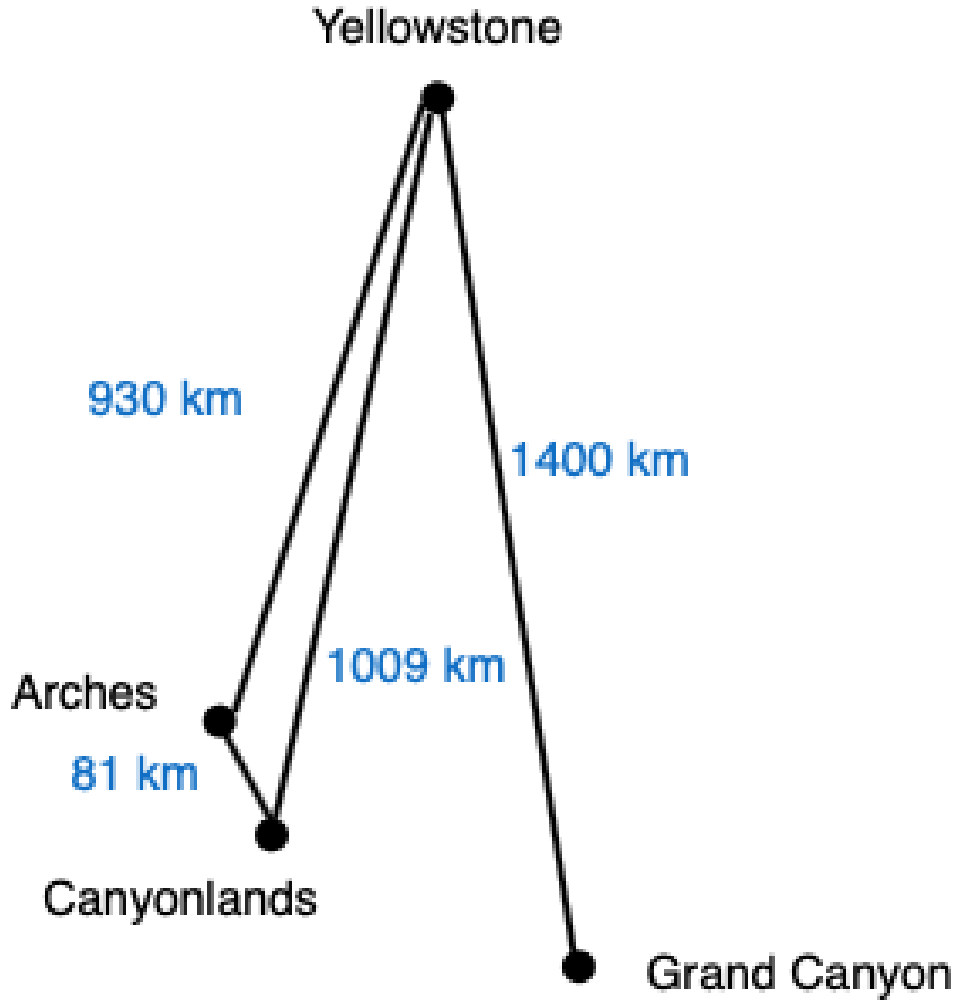
How nMDS Works



How nMDS Works



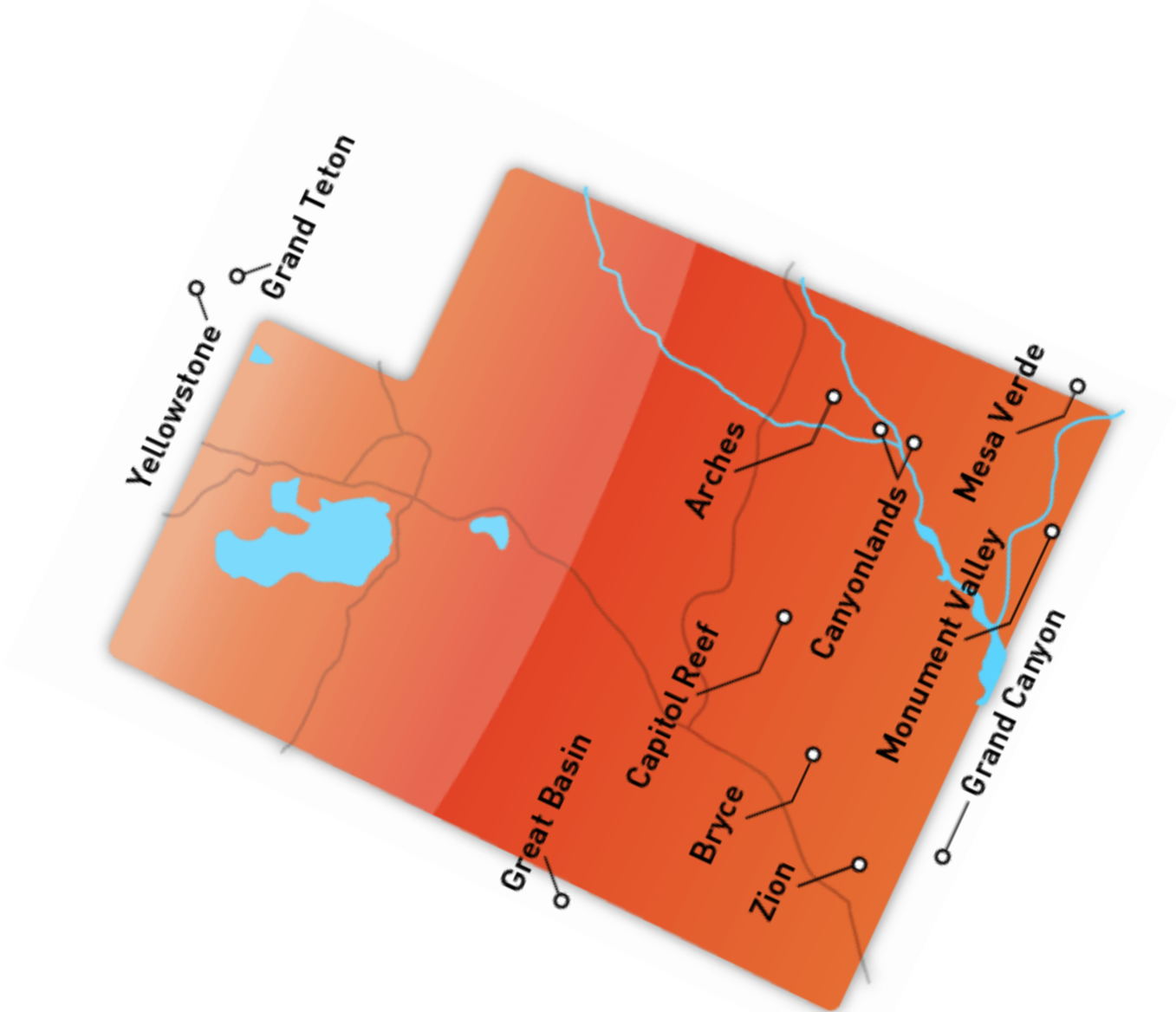
How nMDS Works



How nMDS Works

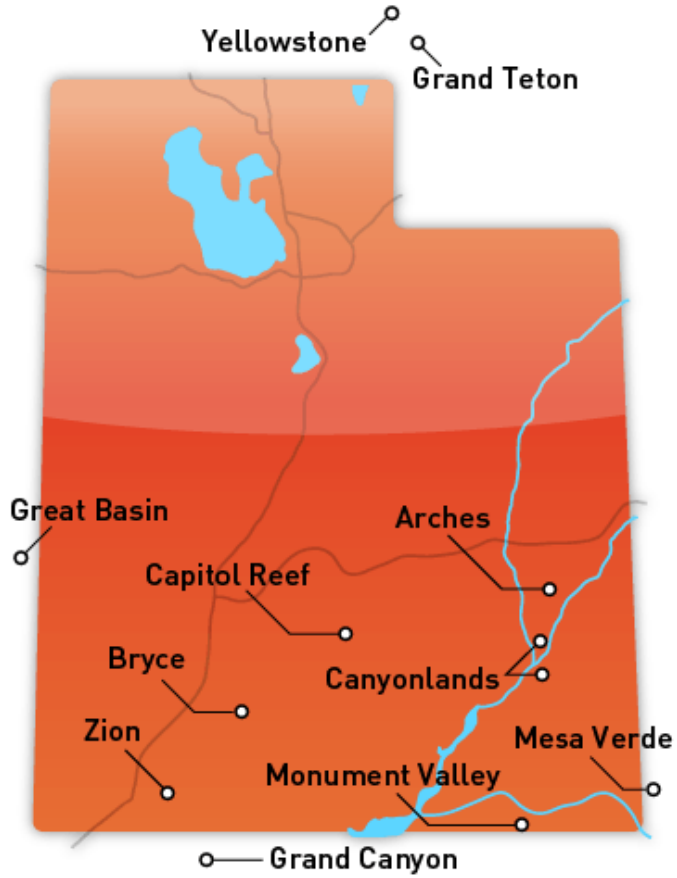
How nMDS Works

No axes in nMDS, so you can flip images. PCA has defined axes so you can't flip them.



How nMDS Works

Note: points that are close together have a shorter distance between them. This is equivalent to sites/samples being more similar and thus having smaller distances (d_{ij}).



The nMDS Process

As I mentioned, nMDS is an iterative process, which occurs over several steps.

1. Define original data points in multidimensional space based on distances between sites (d_{ij}).
2. Specify the number of reduced dimensions you want
   - Typically you shoot for 2 dimensions.
3. Construct an initial configuration of the data.
4. Compare distances in this intial configuration against calculated distances.
5. Determine stress on data points.
6. Correct the position of the points in the dimensional space you have chosen (here 2D) to optimize stress for all points.

nMDS and Stress

Stress: is a value that describes the difference between the distance values from multidimensional space calculated from the distance matrix (d_{ij}), and the distance between points in the reduced dimension representation.

- nMDS tries to optimize stress. "Pulling on all points a little bit so no single point is completely wrong, all points are a little off compared to distances"
- We want to reduce stress as much as we can in reduced dimension. Stress of 5D doesn't help us understand our data well because 5D is hard to understand.

Data Example

How does increasing connectivity of waterways alter aquatic zooplankton community?

Journal of Applied Ecology

BRITISH ECOLOGICAL SOCIETY

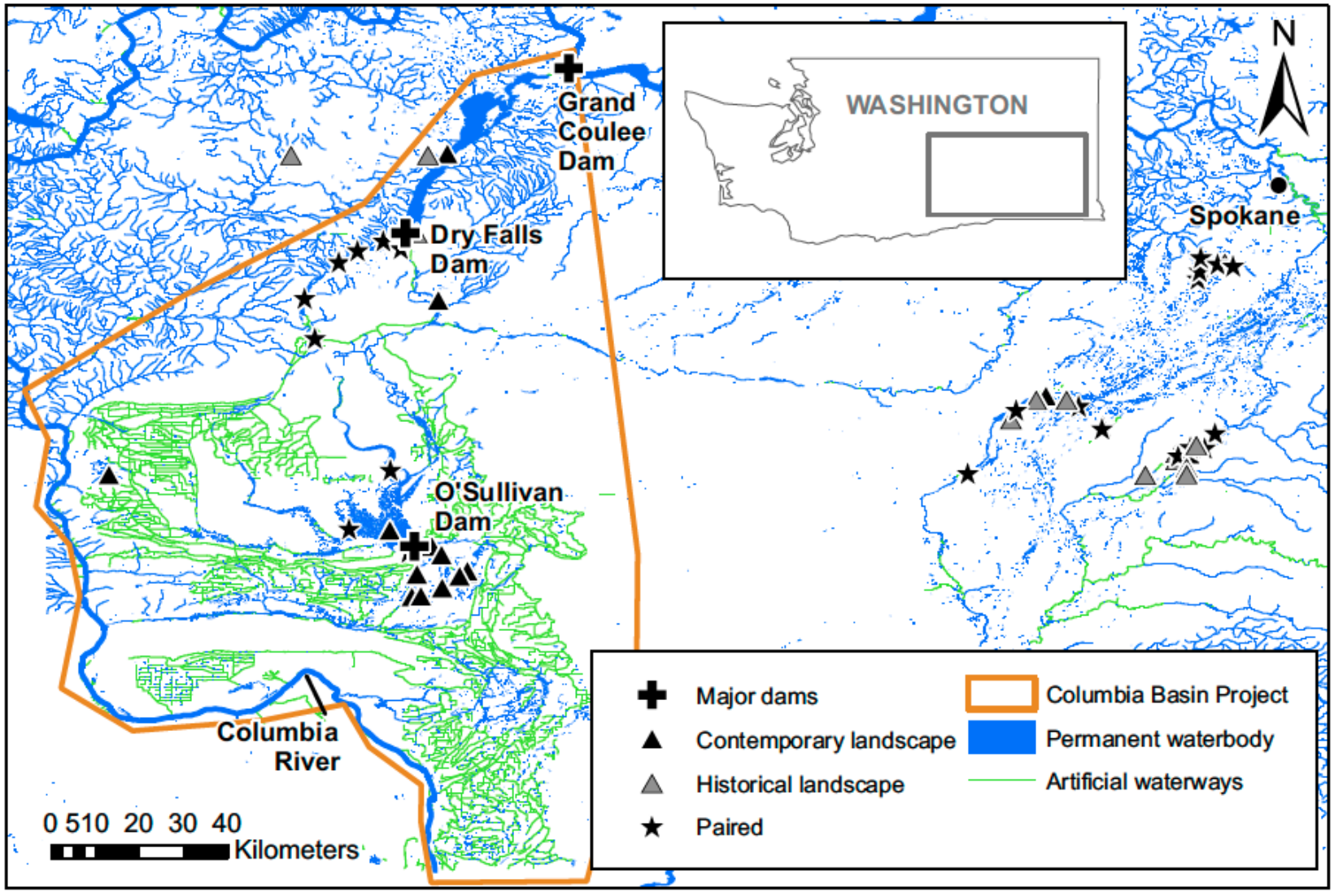
Journal of Applied Ecology 2017, **54**, 1343–1352 doi: 10.1111/1365-2664.12882

Increased habitat connectivity homogenizes freshwater communities: historical and landscape perspectives

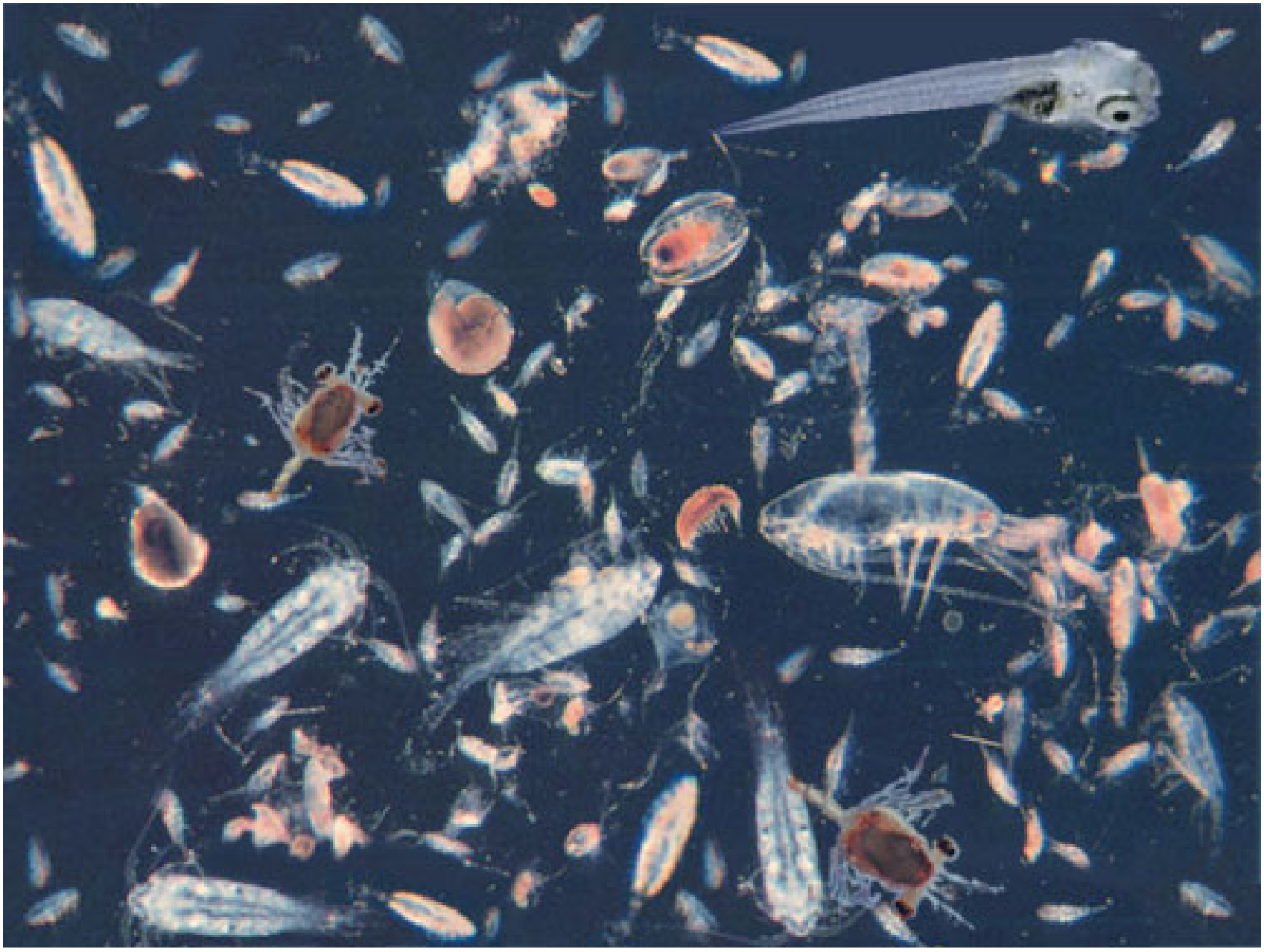
Angela L. Strecker* and Jeffrey T. Brittain

Department of Environmental Science and Management, Portland State University, Portland, OR, USA

Data Example



Data Example



Zooplankton Community Data

Abundance data are normalized by site totals already.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Acanthocyclc	Acanthocyclc	Aglaodiaptor	Bosmina.cor	Bosmina.spp	Camptocercu	Ceriodaphnia	Ceriodaphnia	Ceriodaphnia	Chydoridae.s	Chydorus.spp
2	Alkali	0.06641382	0	0	0	0	0	0	0	0.43607565	0.00501236	0
3	Banks	0	0	0	0.1797235	0.02304148	0	0	0	0	0	0
4	BillyClapp	0	0	0	0.12834225	0	0	0	0	0	0	0.02139037
5	Blue	0	0	0	0	0	0	0	0	0	0	0
6	Canal1	0	0	0	0.31434051	0.13469349	0	0	0	0.00183825	0.17095712	0.11213317
7	Canal2	0	0	0	0.13401875	0.12886418	0	0.00257728	0	0	0	0.10566863
8	CanalLake	0	0	0.02684564	0	0.01118568	0	0	0	0.06823266	0	0.00335571
9	Clear	0	0	0	0	0.0021645	0	0	0	0	0	0
10	ClearPot	0	0	0.01221996	0	0	0	0	0	0.13034623	0.00610998	0.33401222
11	ClearSprague	0	0.20673077	0	0	0	0	0.59615385	0	0	0	0.11538462
12	Cow	0	0	0	0	0.19965189	0	0.74178356	0	0	0	0
13	Deep	0	0	0	0	0	0	0	0	0.07208505	0	0
14	DryFalls	0	0	0.18351477	0	0	0	0.08942457	0	0	0	0.00388803
15	EastTwin	0	0	0.0251454	0	0	0	0	0	0	0.47107845	0.254637
16	Evergreen	0	0	0	0.0952381	0.16666667	0.01428571	0	0	0	0	0.06666667
17	Halfmoon	0	0	0	0	0.06788866	0	0.06279701	0.05261371	0	0	0.0203666
18	Heart	0	0	0	0	0.00673377	0	0	0	0	0	0
19	Hutchinson	0	0	0	0	0.35200523	0	0.31168265	0	0	0	0
20	Lenore	0	0	0	0	0	0	0	0	0.13963039	0	0
21	LilyPad	0	0	0.00382979	0	0	0	0	0	0.36178049	0	0.10191
22	LowerGoose	0	0	0	0.06060606	0	0	0	0	0	0	0
23	Marsh	0	0	0.03897181	0	0	0	0.93864013	0	0	0	0.00165838
24	Medical	0	0	0.03067485	0	0.01431493	0	0	0	0	0	0
25	Miller	0	0.60473588	0	0	0	0	0	0	0.13843352	0	0.16757741
26	Morgan	0	0	0	0	0.46244204	0	0.17908295	0.18928387	0	0	0.00226687
27	Moses	0	0	0	0.25925926	0	0	0	0	0	0	0.04214559
28	NorthTeal	0	0	0	0	0	0	0	0	0	0	0

Running an nMDS

To run an nMDS, use the `metaMDS()` function in `library(vegan)`.

- By default it will include 20 random starts, that's good!
- You can specify the distance matrix you want with `distance=` (default is Bray-Curtis).
- The function also allows you to transform your data, and expand on species scores, but you want to do that yourself beforehand with decostand so you know what you are doing.
  - Thus, `autotransform=` and `expand=` should mostly always be FALSE
  - Note, as I mentioned, the data have already been transformed by site totals in our current case.

Note: when you run an nMDS you want to remove all columns that are not species abundances (e.g. remove site columns, any other treatment or environmental variables, etc.)

Running an nMDS

```
zoop.mds2 <- metaMDS(zoop[,-1], k = 2, distance = 'bray', autotransform = F,
                     expand = F)
```

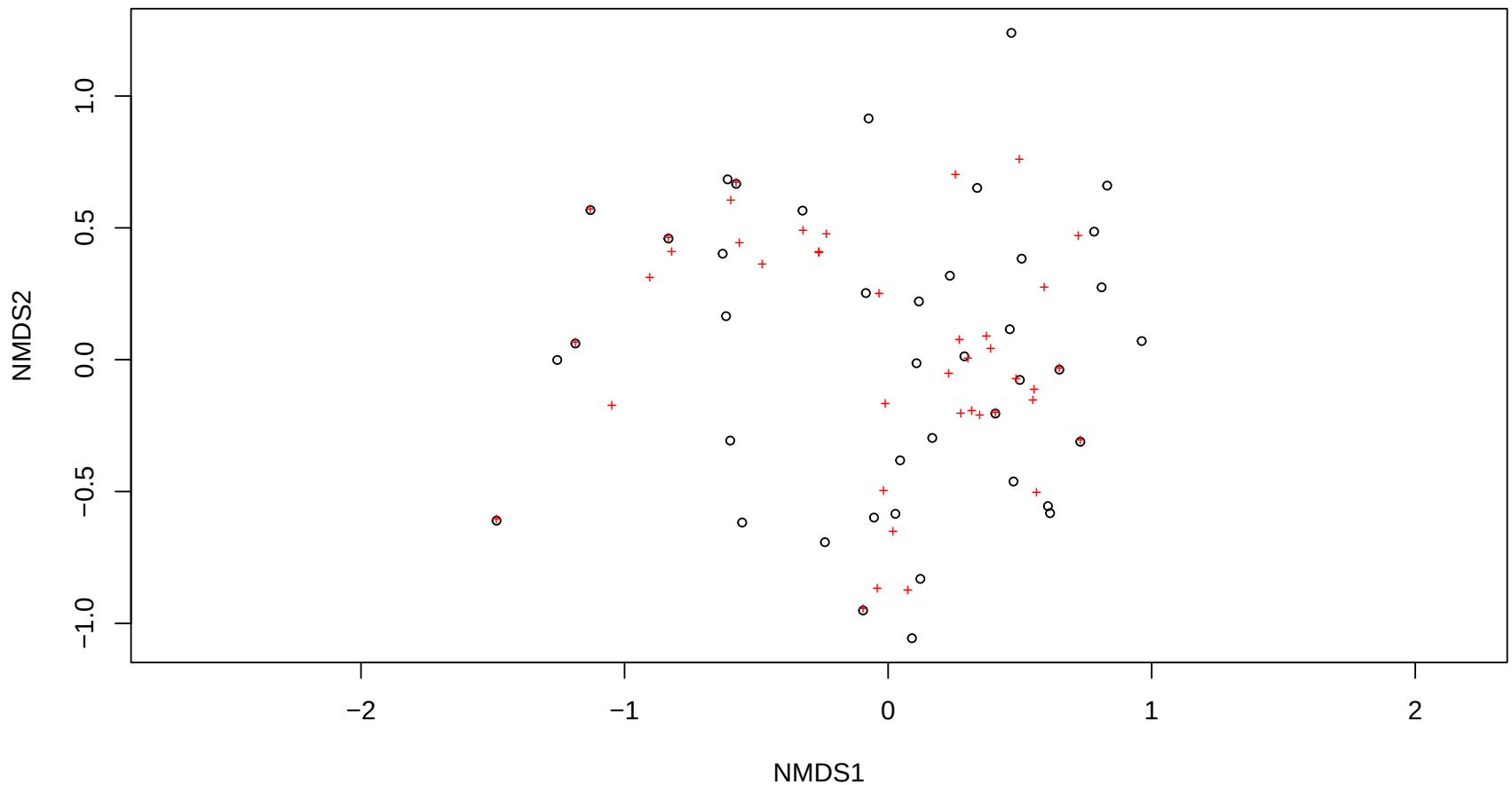
```
## Run 0 stress 0.1782076
## Run 1 stress 0.1781714
## ... New best solution
## ... Procrustes: rmse 0.008741075  max resid 0.04268711
## Run 2 stress 0.1781715
## ... Procrustes: rmse 0.0001073504  max resid 0.0004695505
## ... Similar to previous best
## Run 3 stress 0.1781715
## ... Procrustes: rmse 8.170931e-05  max resid 0.0003501646
## ... Similar to previous best
## Run 4 stress 0.2028361
## Run 5 stress 0.1782076
## ... Procrustes: rmse 0.008741628  max resid 0.04285159
## Run 6 stress 0.1781714
## ... Procrustes: rmse 6.879909e-05  max resid 0.0002860371
## ... Similar to previous best
## Run 7 stress 0.202642
## Run 8 stress 0.2026417
## Run 9 stress 0.2028783
## Run 10 stress 0.1781714
## ... Procrustes: rmse 5.145087e-05  max resid 0.0002312016
## ... Similar to previous best
```

Plotting an nMDS

This is the biplot - it shows sites (circles), and species (red +).

- *Remember:* points close together in space are more similar, and points far away are more different.

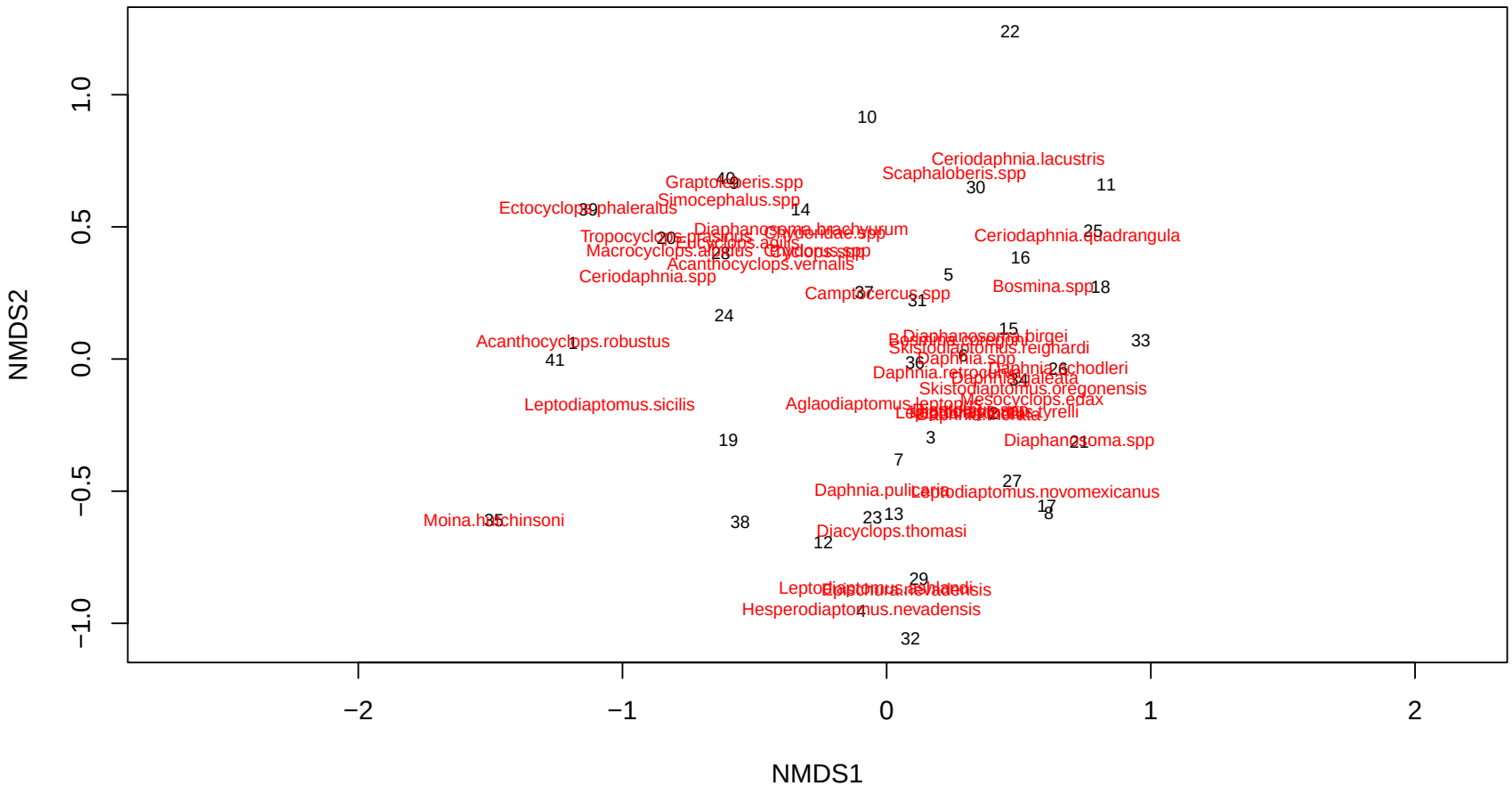
```
plot(zoop.mds2)
```



Plotting an nMDS

Here, `type = "t"` where "t" = text and shows the site and species names.

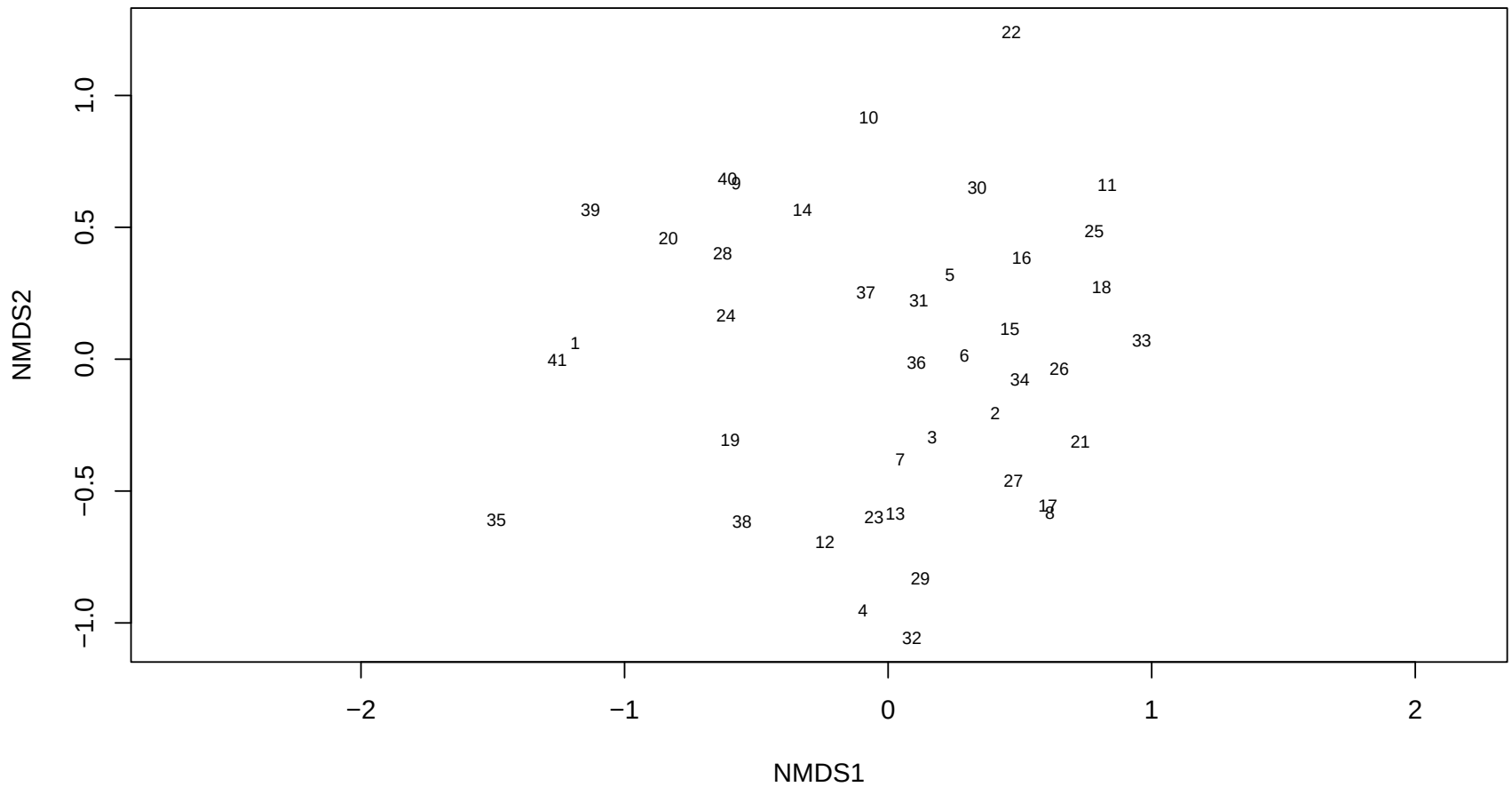
```
plot(zoop.mds2, type = "t")
```



Plotting an nMDS

If you just want to see the site names use `disp = "sites"`. You can also try `disp = "species"`.

```
plot(zoop.mds2, disp = "sites", type = "t")
```



Dimension vs Stress

nMDS is trying to solve a multidimensional problem and represent it in a given number of dimensions (`k`). As you increase your dimension, you will decrease your stress, but increased dimension also makes it more difficult to interpret the cloud of points (`k=2` is understandable because it's 2D).

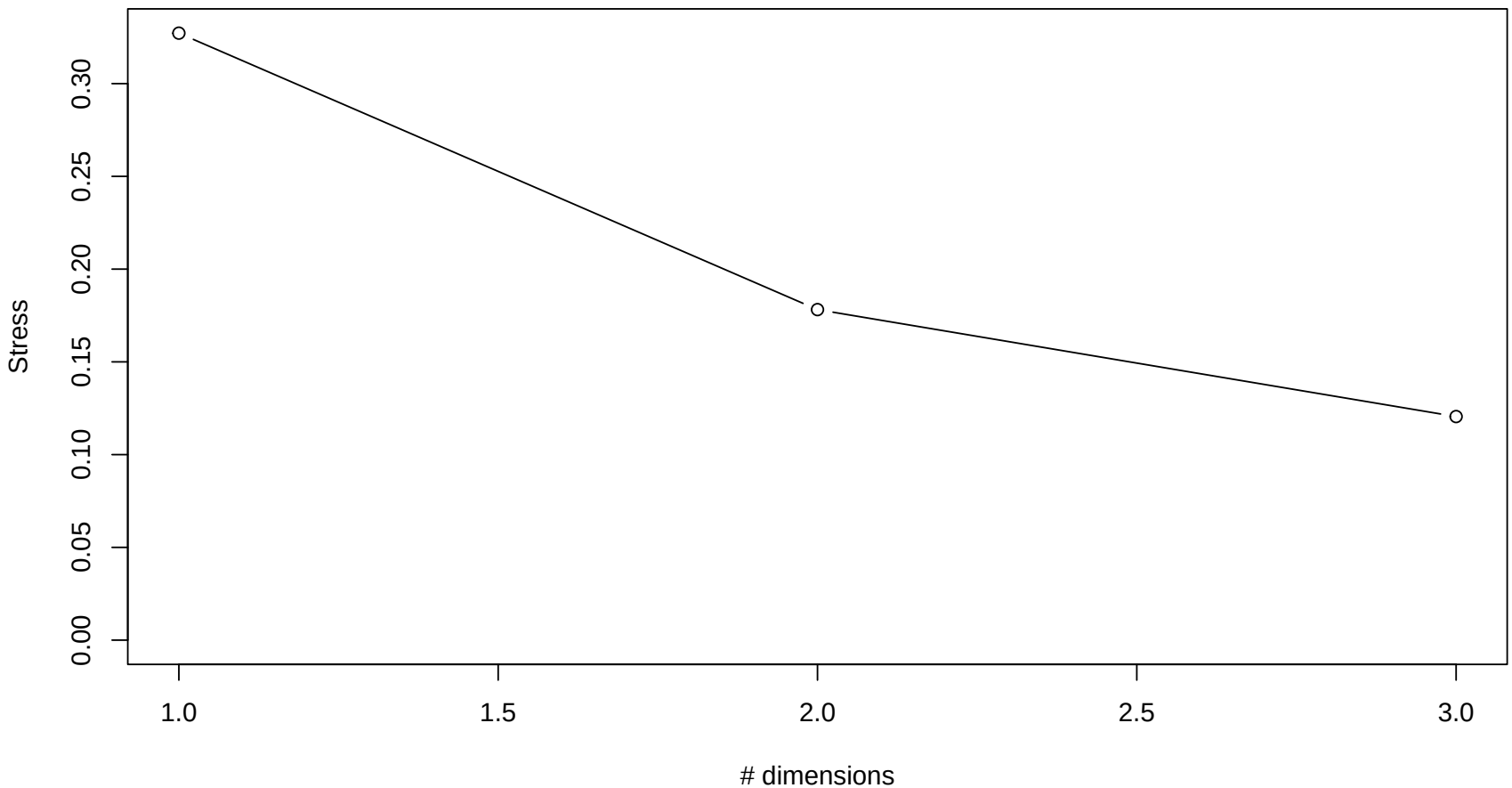
- You want to optimize the number of dimensions vs the amount of stress.

```
zoop.mds1 <- metaMDS(zoop[,-1], k = 1, distance = 'bray', autotransform = F,
                     expand = F)
zoop.mds3 <- metaMDS(zoop[,-1], k = 3, distance = 'bray', autotransform = F,
                     expand = F)
```

Dimension vs Stress

You can plot the number of dimensions vs stress to make an informed decision.

- <10% stress is ideal, but that's often not possible, so report your dimension and stress and let readers interpret.



Adding Environmental Data

You can next find how your environmental variables for all sites plot along the nMDS cloud of points.

To do this, you **must** make sure your data are in the same order. To me, the easiest way to do this is to merge the composition and environmental datasets and then remove columns when you are running various anlaysis.

```
env[1:8, 1:6]
```

```
## # A tibble: 8 x 6
##   site       category  type      surface_area surface_temp elevation
##   <chr>      <chr>     <chr>            <dbl>        <dbl>     <dbl>
## 1 Alkali     landscape pond              9.7          26.8       568
## 2 Banks      landscape reservoir     10926.           20.2       479
## 3 BillyClapp landscape reservoir       405.           20.7       407
## 4 Blue       paired    lake            214.           24.3       335
## 5 Canal1     landscape canal            NA            21.8       305
## 6 Canal2     landscape canal            NA            21.3       308
## 7 CanalLake  landscape lake             24.7          22.9       300
## 8 Clear      paired    lake            166.           24         713
```

Merge Environmental Data and rerun nMDS

```
dat <- full_join(zoop, env, by = "site")
dat[1:10,-c(3:45)]
```

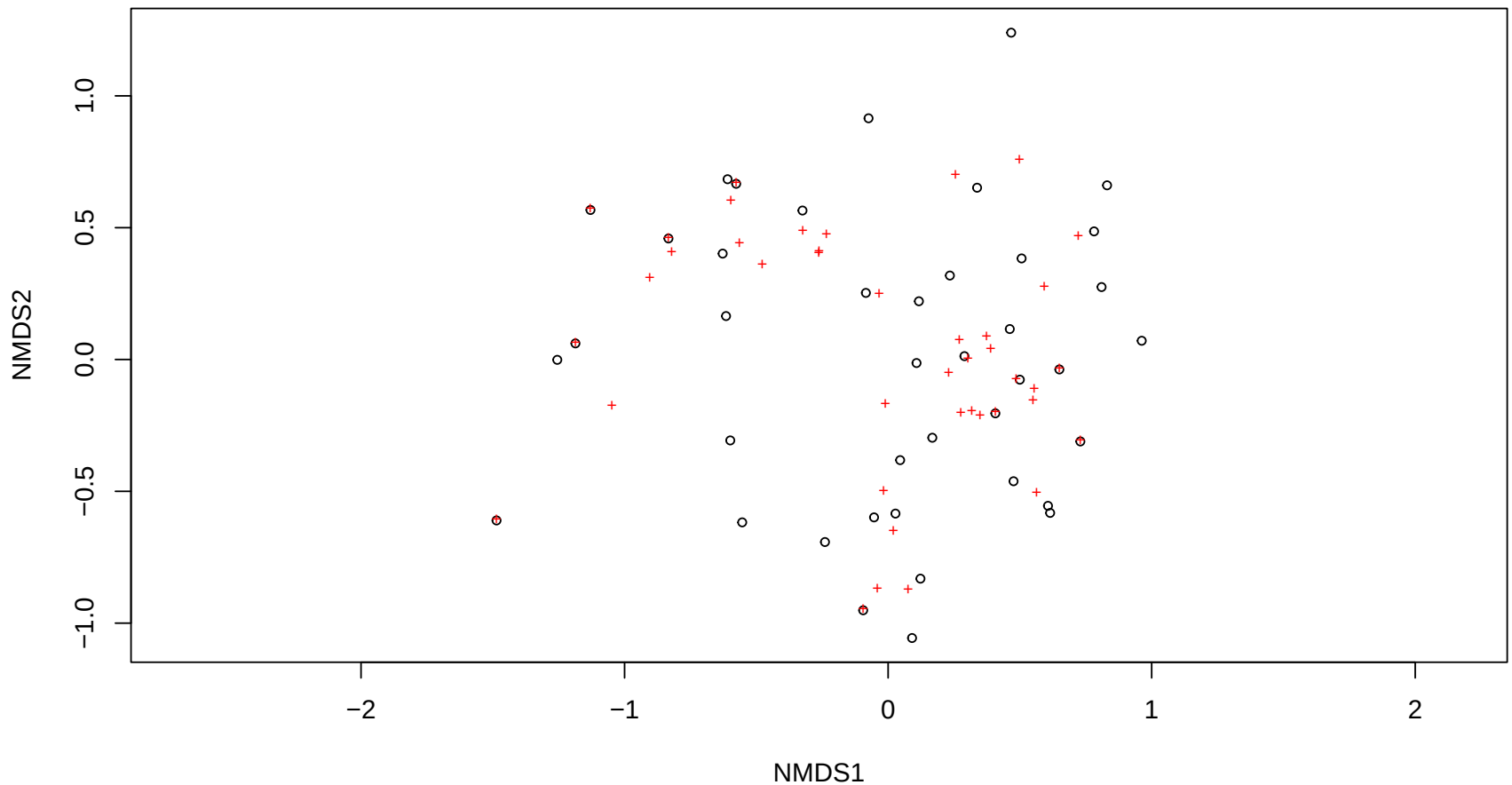
```
## # A tibble: 10 x 5
##    site         Acanthocyclops.robustus surface_temp elevation secchi
##    <chr>                          <dbl>        <dbl>     <dbl>  <dbl>
##  1 Alkali                        0.0664         26.8       568    0.5
##  2 Banks                         0              20.2       479    2.5
##  3 BillyClapp                    0              20.7       407    4.8
##  4 Blue                          0              24.3       335    4.3
##  5 Canal1                        0              21.8       305    1
##  6 Canal2                        0              21.3       308   NA
##  7 CanalLake                     0              22.9       300    4.9
##  8 Clear                         0              24         713    3
##  9 ClearPot                      0              20.5       719    1.5
## 10 ClearSprague                  0              23.2       592    0.4
```

nMDS with Environmental Data

Run the nMDS with "dat", but remove the site names and environmental variables from this dataset.

```
dat.mds2 <- metaMDS(dat[,-c(1, 43:48)], k = 2, distance = 'bray',
                    autotransform = F, expand = F)
```

```
plot(dat.mds2)
```



nMDS with Environmental Data

Create an environemntal dataset that's in the same site order as the composition dataset. FYI: You still don't want plot names.

```
dat.env <- dat[,43:48]
dat.env
```

```
## # A tibble: 41 x 6
##    category  type      surface_area surface_temp elevation secchi
##    <chr>     <chr>            <dbl>        <dbl>     <dbl>  <dbl>
##  1 landscape pond               9.7          26.8       568    0.5
##  2 landscape reservoir      10926.           20.2       479    2.5
##  3 landscape reservoir        405.           20.7       407    4.8
##  4 paired    lake             214.           24.3       335    4.3
##  5 landscape canal             NA            21.8       305    1
##  6 landscape canal             NA            21.3       308   NA
##  7 landscape lake              24.7          22.9       300    4.9
##  8 paired    lake             166.           24         713    3
##  9 paired    pond               0.3          20.5       719    1.5
## 10 landscape pond               2.4          23.2       592    0.4
## # ... with 31 more rows
```

nMDS with Environmental Data

Create an environemntal fit to the nMDS data with `envfit()` in `library(vegan)`, and add the environmental variables to the nMDS plot.

- Not all of this environmental fit code works with non-vegan ordinations.

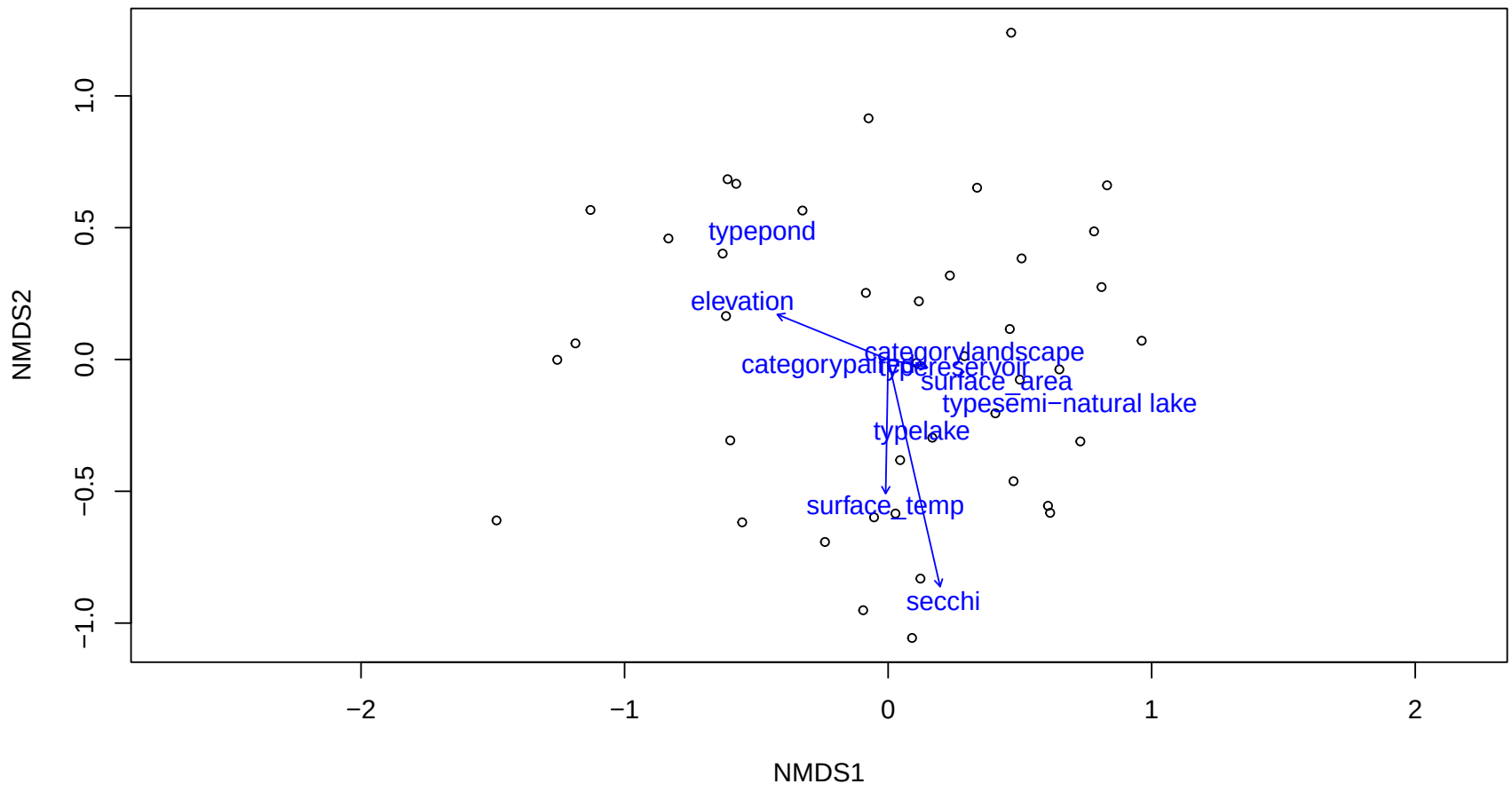
```
dat.efit <- envfit(dat.mds2, dat.env, na.rm = TRUE)

# Here you need na.rm = TRUE because not all sites have values for everything.
```

nMDS with Environmental Data

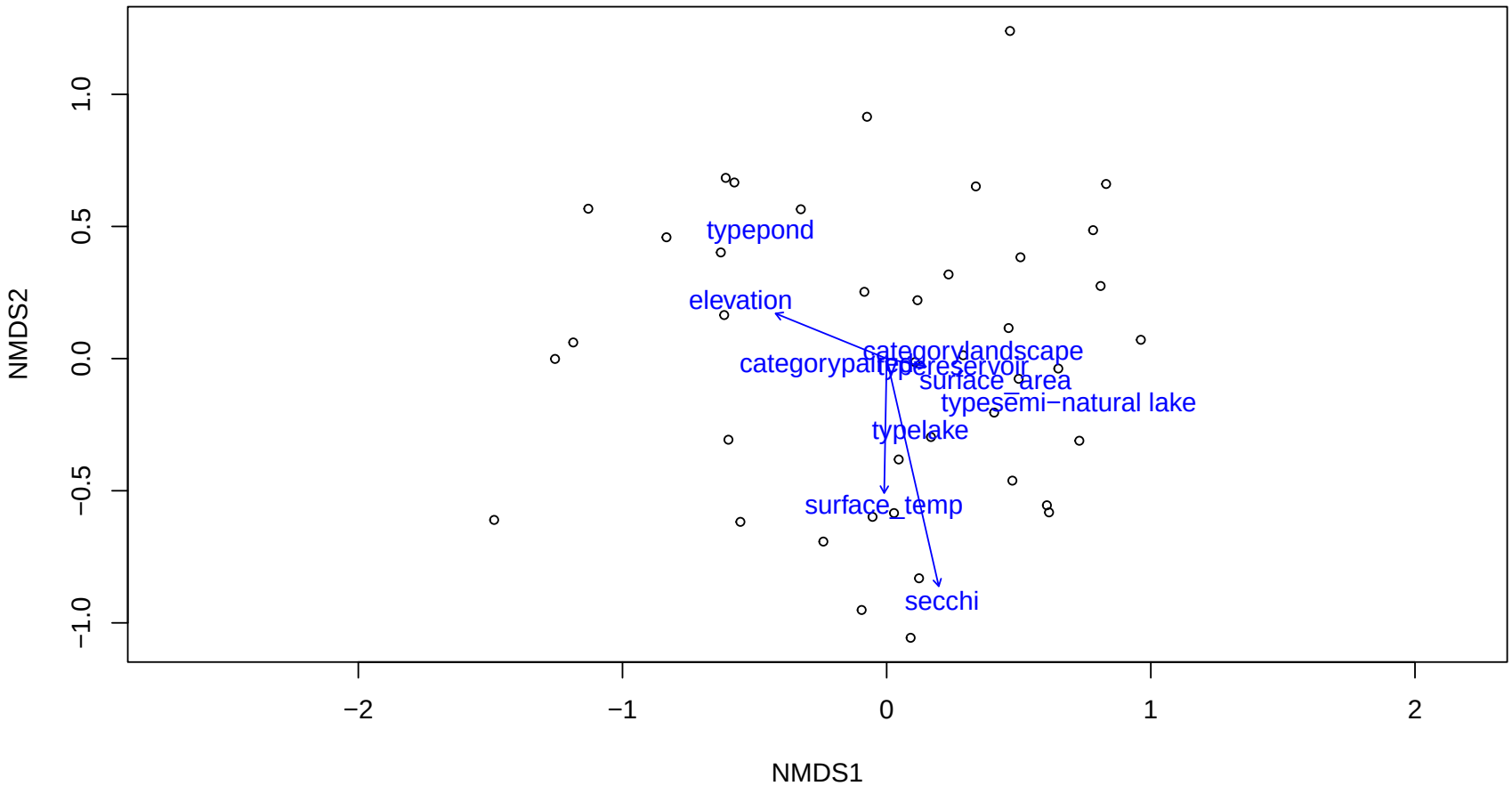
This is a triplot (has info about sites, species, and environment).

```
plot(dat.mds2, disp = "sites")
plot(dat.efit)
```



Interpreting Triplots

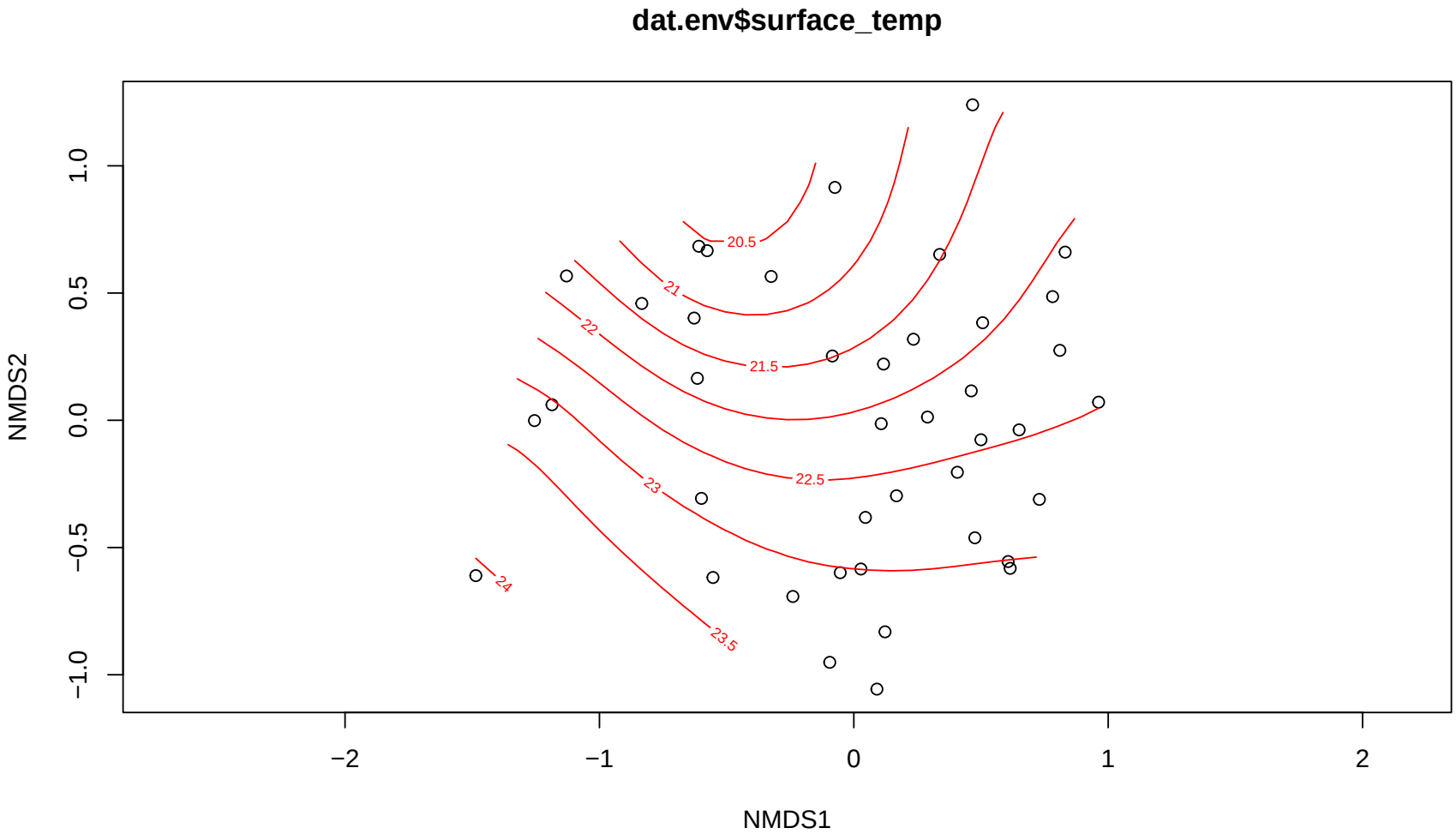
- Direction of arrows indicates the trend of sites (toward the arrow indicates more of the variable).
- Length of the arrow indiciates stronger relationship (**but this is not a statistical test!**)



Fun with Triplots

Draw a surface for an environmental variable of interest.

```
ordisurf(dat.mds2, dat.env$surface_temp)
```

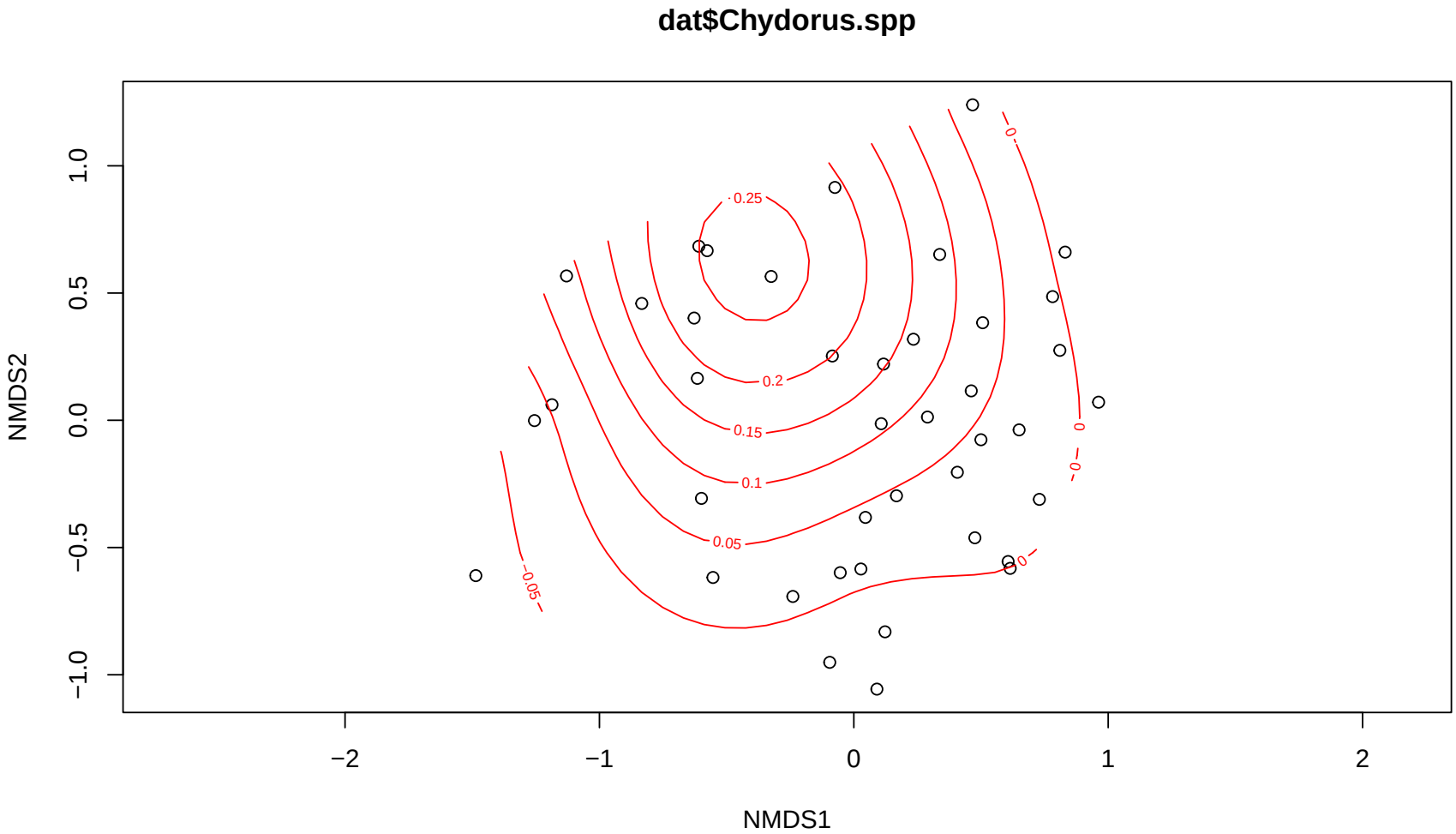


##

Fun with Triplots

Draw a surface for a species of interest.

```
ordisurf(dat.mds2, dat$Chydorus.spp)
```

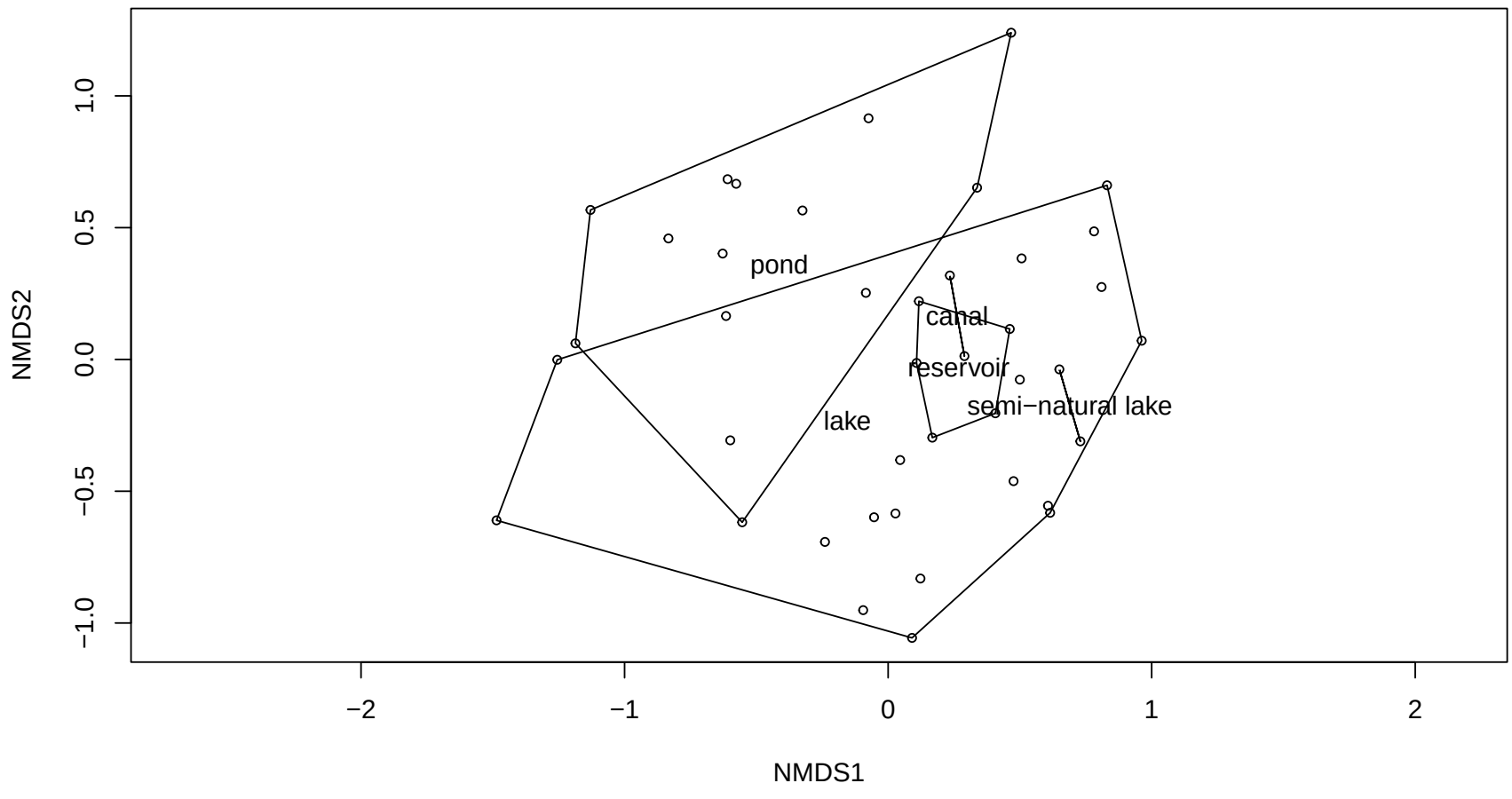


##

Fun with Triplots

Add convex hulls for groups, added on top of nMDS plots.

```
plot(dat.mds2, disp = "sites")
ordihull(dat.mds2, dat.env$type, label = T)
```



Fun with Triplots

Draw spider diagrams connecting sites to group centroids.

```
plot(dat.mds2, disp = "sites")
ordispider(dat.mds2, dat.env$type, label = T)
```