# Metapopulation Genetics

Helene Wagner, University of Toronto

D. depressissimum

Goal: Test genetic effects of metapopulation dynamics

- Quantify spatial genetic structure
- Test effects of population size and connectivity
- Test effect of recent extinction events

Photos: J-P Pointier

Methodological Challenges

1. 'EcoGenetics': multi-source data

2. AMOVA with R

3. Regression analysis in R:
    - Interpret results
    - Check assumptions

4. Video 2: graphics with 'ggplot2'

Freshwater Snail Data

Lamy et al. (2012), Molecular Ecology 21: 1394 – 1410

Genetic Diversity of *D. depressissimum*

25 sites total:
- 12 sites: 4 clusters
- 12 sites: multi-year

Allelic Richness (rarefied)

Genetic data:
- 22-32 individuals per site
- 10 microsatellites

Population-level:
- RA: allelic richness
- He: expected heterozygosity
- FST.GESTE: site-level FST
- Fst.temp: temporal FST

Predictor variables:
- NLT: long-term pop size
- Size: pond size
- C: hydrological connectivity
- D: wetland area within 2 km
- APE: apparent extinction event?
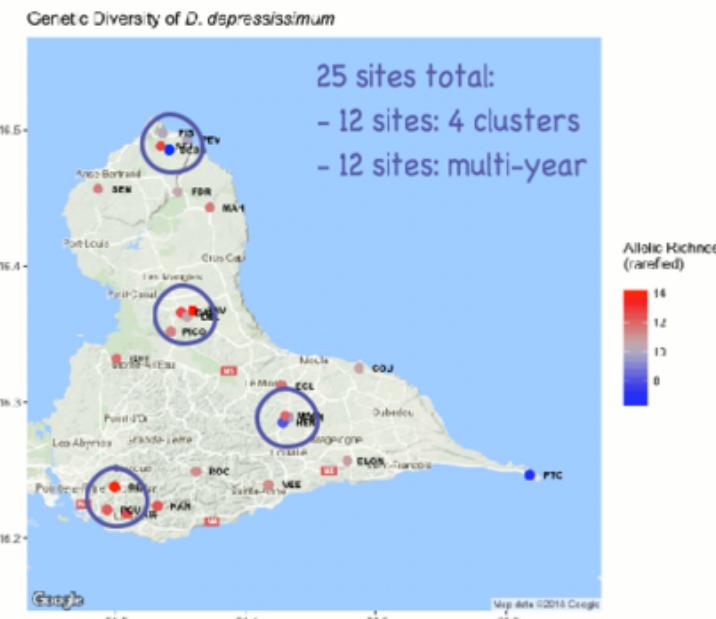
Hierarchical data set

4 clusters x 3 ponds = 12

AMOVA:
Differentiation within vs. between clusters

Within < Between

Spatial data set

25 sites (out of 2059 ponds)

Multiple regression:
RA ~ NLT + C
FST.GESTE ~ NLT + C

Diversity increases with NLT
Differentiation decreases with C

Temporal data set

12 ponds x 2-4 years
(5 with APE, 7 without)

Two-sample t-test:
Fst.temp ~ APE

Higher Fst.temp if exctinct

Package 'EcoGenetics'

|| ECOGEN CLASS OBJECT ||

```
ecoslot.G(obj)       Access to slots: <ecoslot.> + <name of the slot> + <(name of the object)>
obj[["G"]]           See: help("EcoGenetics accessors")
Not: obj @ G
```

|			
Spatial coordinates	| slot XY:	| --> 1270 x 2	coordinates
Phentypic traits	| slot P:	| --> 0 x 0	phenotypic variables
Genetic data	| slot G:	| --> 1270 x 10	loci >> ploidy: 2 || codominant
Allele frequencies	| slot A:	| --> 1270 x 372	alleles
Site variables	| slot E:	| --> 0 x 0	environmental variables
Strata (sampling)	| slot S:	| --> 1270 x 4	structures >> 4 structures found
Anything else?	| slot C:	| --> 0 x 0	variables
Result output	| slot OUT:	| --> 0	results

SiteID, SITE
Cluster, YEAR

Why use 'EcoGenetics'?

1. Integrate multi-source data
2. Export / import, e.g.: 'ecogen2genind'
3. Calculate spatial statistics (Week 5)

Data Management:

– Slicing: obj[1:3,]
– Constructing 'ecogen': row names must match
– Can join by column: 'eco.fill_ecogen_with_df'
– Aggregate to 'ecopop' object: 'ecogen2ecopop'

Hierarchical Levels

Cluster   Cluster

Pond   Pond

Snail   Snail

AMOVA

A-nalysis of
MO-lecular
VA-riance
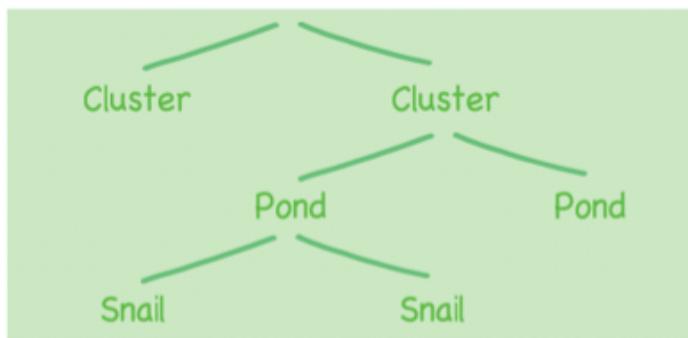
Why?

FST:
- Genetic structure relative to fixation
- Parameters rely on pop gen assumptions

AMOVA:
- No underlying pop gen model

```
> amova.result
$call
ade4::amova(samples = xtab, distances = xdist, structures = xstruct)

$results
                                 Df    Sum Sq    Mean Sq
Between Cluster                   3   106.2413  35.413781
Between samples Within Cluster    8   249.6422  31.205270
Within samples                  353  2764.2124   7.830630
Total                           364  3120.0959   8.571692

$componentsofcovariance
                                             Sigma           %
Variations  Between Cluster               0.04446256   0.5142958    a
Variations  Between samples Within Cluster 0.77023554   8.9092691    b
Variations  Within samples                7.83062999  90.5764351    c
Total variations                          8.64532809 100.0000000

> amova.test
 class: krandtest lightkrandtest
 Monte-Carlo tests
 Call: randtest.amova(xtest = amova.result, nrepet = 999)

 Number of tests:    3

 Adjustment method for multiple comparisons:    none
 Permutation number:    999
                              Test      Obs     Std.Obs   Alter  Pvalue
 1  Variations within samples 7.83062999 -80.175606    less   0.001
 2  Variations between samples 0.77023554  57.660752 greater   0.001
 3  Variations between Cluster 0.04446256   1.781495 greater   0.050
```

Interpretation

1. Variance explained at each level?
2. Significantly > 0?
3. Phi = population differentiation statistics

```
$statphi
                         Phi
Phi-samples-total    0.094235649    Phi.ST = a + b
Phi-samples-Cluster  0.089553260    Phi.SC = b/(b+c)
Phi-Cluster-total    0.005142958    Phi.CT = a
```

Regression Interpretation

```
lm ( FST.GESTE ~ NLT + C, data = dd.site )
```

Cohen's Effect Size

	Correlation	Regression
S	r > 0.1	$R^2 > 0.01$
M	r > 0.3	$R^2 > 0.09$
L	r > 0.5	$R^2 > 0.25$

Direction?

Sign of slope

As expected?

Effect Size?

Multiple R-squared
Beta coefficients

Biologically relevant?

Significant?

p-values of t-tests

Rule out chance?
Correlation is not causation!

Variation Partitioning

NLT   C

0.26   0   0.20

Unexplained: 0.54
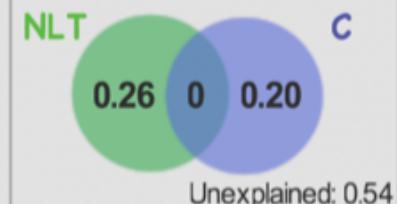
With Original Variables

```
Coefficients:  Slopes                          t-Tests
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.197237   0.034235   5.761 8.53e-06 ***
NLT         -0.042685   0.012774  -3.342  0.00296 **
C           -0.030047   0.009946  -3.021  0.00628 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04187 on 22 degrees of freedom
Multiple R-squared:  0.4599,    Adjusted R-squared:  0.4108
F-statistic: 9.367 on 2 and 22 DF,  p-value: 0.00114   AdjR2
```

All Variables Standardized

```
Coefficients:  Beta coefs
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.576e-17  1.535e-01   0.000  1.00000
scale(NLT)  -5.254e-01  1.572e-01  -3.342  0.00296 **
scale(C)    -4.750e-01  1.572e-01  -3.021  0.00628 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7676 on 22 degrees of freedom
Multiple R-squared:  0.4599,    Adjusted R-squared:  0.4108
F-statistic: 9.367 on 2 and 22 DF,  p-value: 0.00114
```

Checking Assumptions

OLS Assumptions

Sample:
- Simple random sample
- Randomized experiment

Response Y:
- Linear relationship with X
- Errors normally distributed
- Errors with constant variance
- Errors identically distributed
- Errors independent

Predictors X:
- Measured without error
- No multicollinearity
  - strict: no redundant vars
  - wide: low cor among X

## Variance Inflation Factor (VIF)

Linear relationship with X:
Red line should be horizontal

Residuals vs Fitted

Errors normally distributed:
Points should follow straight line
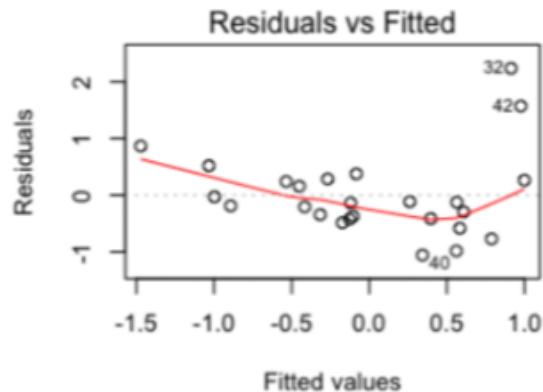
Normal Q-Q

Errors with constant variance:
Red line should be horizontal

Scale-Location
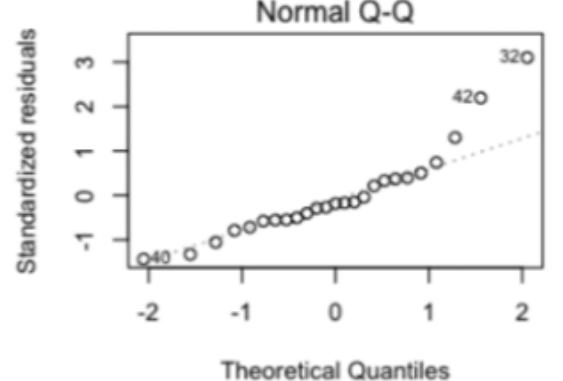
No influential points:
No point with Cook's D > 1

Residuals vs Leverage

Cook's distance

Interpretation:
- Inflation of slope variance
- SE (slope) * sqrt (VIF)
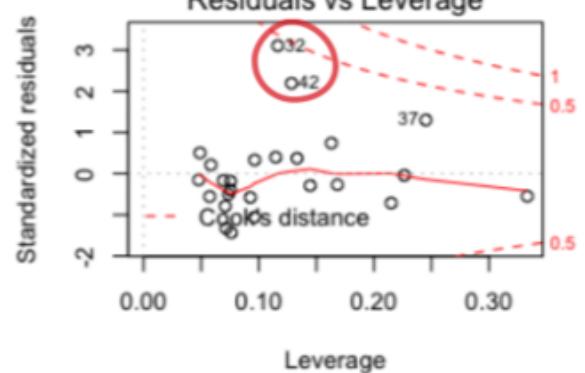
Rule of thumb:
- < 5: moderate collinearity
- > 10: problematic!

The R Roller Coaster

Elation

Full control
Print quality
Reproducible
Amazing features

Steep learning curve
Why not by menu?
Program every detail
Forgetting details

Frustration
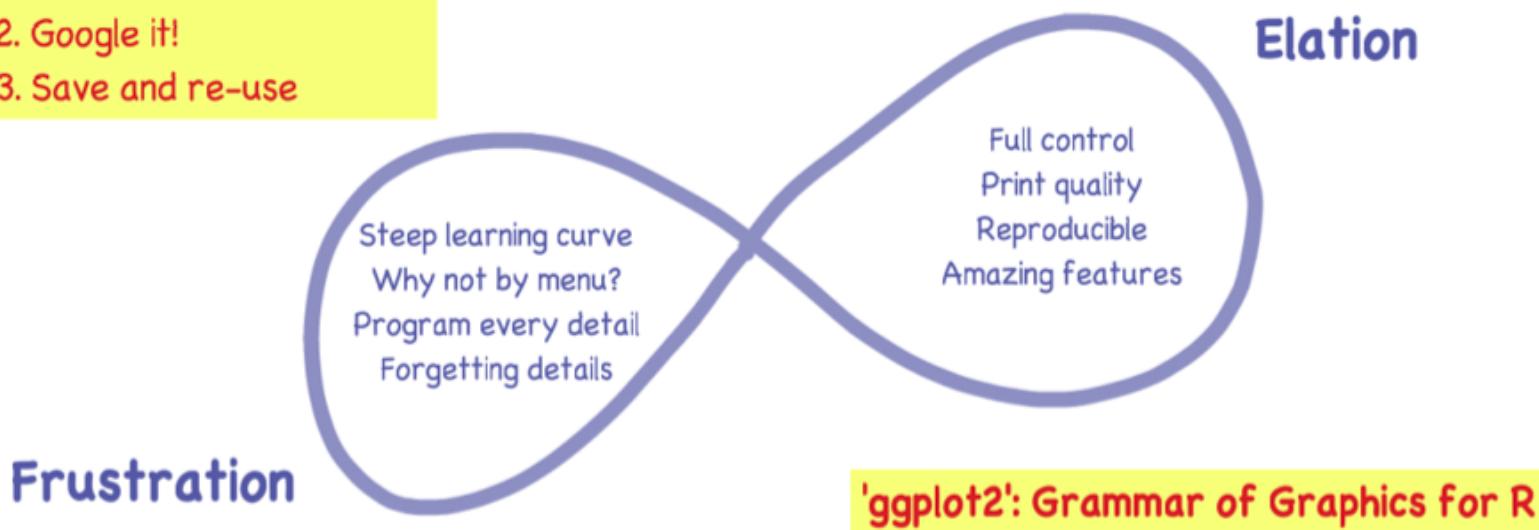
'ggplot2': Grammar of Graphics for R

'lattice': Trellis graphics for R

R base graphics

R
↑
1995

RStudio
↑
2010

1976   1980   1988   1998
S1   S2   S3   S4
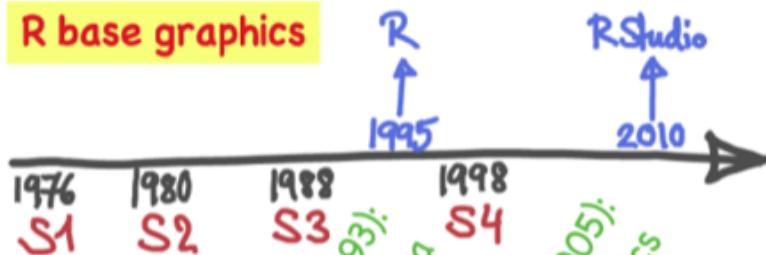
Cleveland (1993):
Visualizing Data

Wilkinson (1999, 2005):
Grammar of Graphics

Hadley Wickham
Chief Scientist at RStudio

Basic R Graphics

File types
- png: for import
- pdf: for printing
- NOT: jpg

Figure margins

side = 3

side = 2   side = 4

side = 1   0
              1 Lines
              2
              3

Site-level Fst — Long-term population size (DESB, PTC)

Site-level Fst — Hydrological connectivity (DESB, PTC)
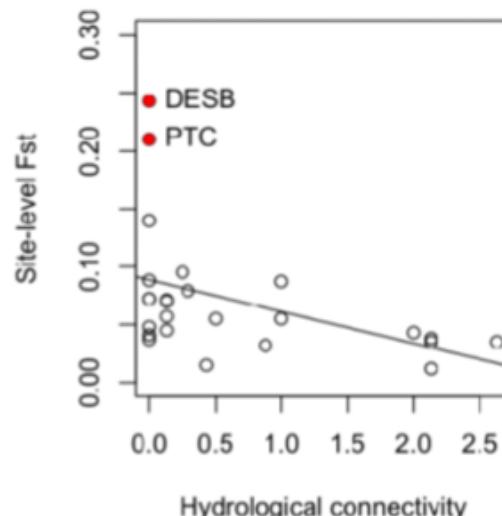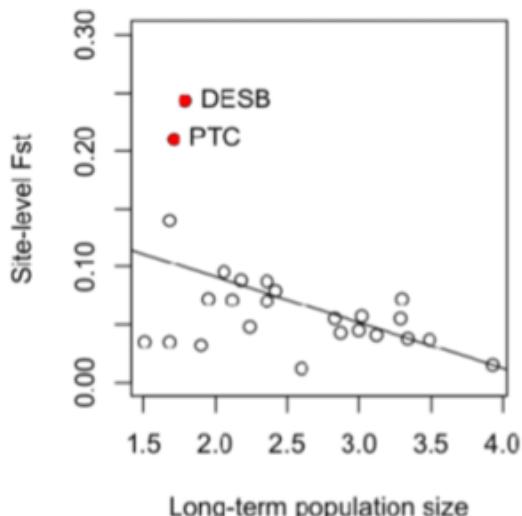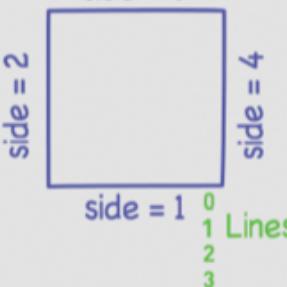
pch	Point Types			lty	Line Types
○ 1	⊠ 14				1
△ 2	■ 15				2
+ 3	● 16				3
× 4	▲ 17				4
◇ 5	◆ 18				5
▽ 6	● 19				6
⊠ 7	● 20			lwd	Line Width
✳ 8	⬡ 21				.1
⊕ 9	⊟ 22				.25
⊕ 10	◇ 23				.5
✴ 11	△ 24				1
⊞ 12	▽ 25				3
⊠ 13	you can also use any character				6

```
png( "MyFigure.png",
    width = 7, height = 3.5, units = "in",
    pointsize = 12, res = 300 )

par ( mfrow = c ( 1, 2 ), mar = c ( 4, 4, 1, 1 ))

  plot ( FST.GESTE ~ NLT, data = df)
  plot ( FST.GESTE ~ C, data = df)
dev.off ( )
```

Define axes
```
plot ( y ~ x, data = df, ylim = c ( 0, 0.3 ), type = "n",
       xlab = "Label x", ylab = "Label y")
```

Add data points
```
points ( y ~ x, data = df, cex = 1)
```

Add fitted line
```
abline ( lm ( y ~ x, data = df ))
```

Create index 'a'
```
a <- is.element ( row.names ( df ), c ("32", "42") )
```

Add red symbols
```
points ( y ~ x, data = df [ a, ], pch = 16, col = "red")
```

Add labels
```
with( df [ a, ], text ( x , y, labels = SITE, pos = 4 ))
```

Grammar of Graphics: 'ggplot2'

See RStudio cheat sheet: Data Visualization with ggplot2

Fundamentals

Faceting: split by group
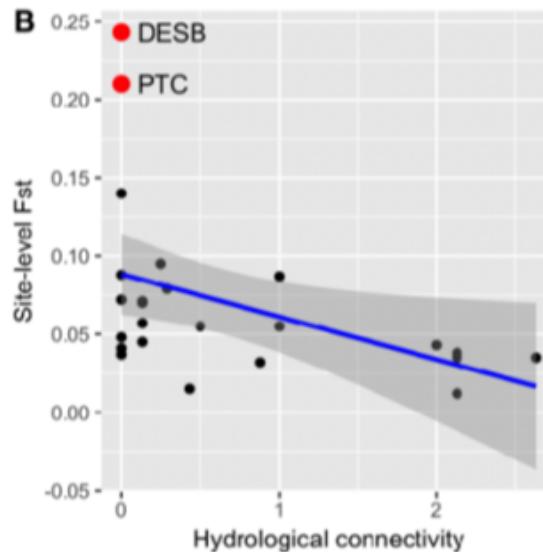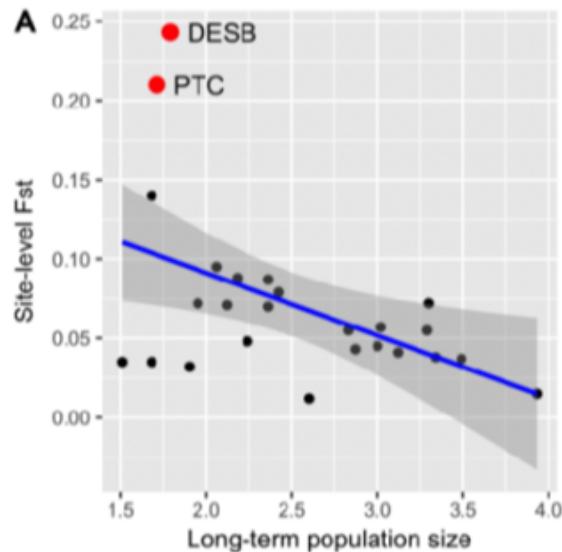
Compose and store plot

Define axes
Add data points
Add fitted line
Add red symbols
Add labels

```
Plot1 <- ggplot(data = df, aes( x, y )) +
    xlab("Label x") + ylab("Label y") +
    geom_point() +
    geom_smooth(col = "blue", method="lm") +
    geom_point(data = df[a,],color = "red",size = 3) +
    geom_text(data = df[a,], mapping=aes(x, y, label = SITE),
            size = 4, hjust = 0, nudge_x = 0.1)
```

Assemble and export figure

```
ggsave(filename = "ggplot.png", width = 7, height = 3.5, unit = "in", dpi = 300,
    plot = cowplot::plot_grid(Plot1, Plot2, labels = c("A", "B")))
```
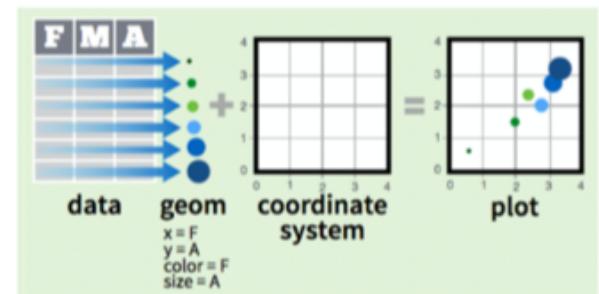
t + **facet_grid(. ~ fl)**
facet into columns based on fl

t + **facet_grid(year ~ .)**
facet into rows based on year

t + **facet_grid(year ~ fl)**
facet into both rows and columns

t + **facet_wrap(~ fl)**
wrap facets into a rectangular layout

Plotting Residuals in Space

```
mod <- lm ( FST.GESTE ~ NLT + C, data = dd.spatial )
```

```
> attributes(mod)
$names
[1] "coefficients"  "residuals"   "effects"  "rank"
[5] "fitted.values" "assign"      "qr"       "df.residual"
[9] "xlevels"       "call"        "terms"    "model"
```

```
Residuals <- mod $ residuals
```

lattice universe: package 'sp'
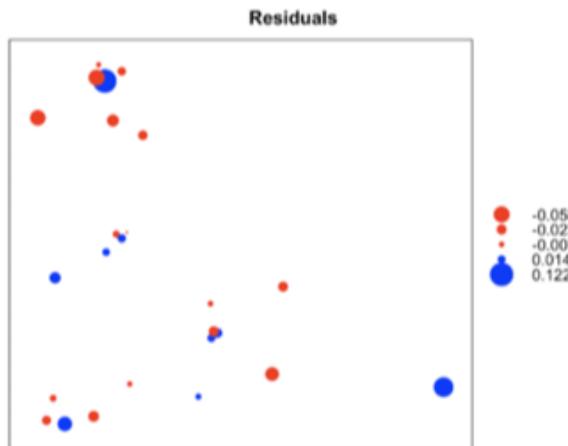

```
sp :: bubble ( dd.spatial, zcol = "Residuals",
          col = c( "red", "blue" ))
```

ggplot2 universe: package 'ggmap'
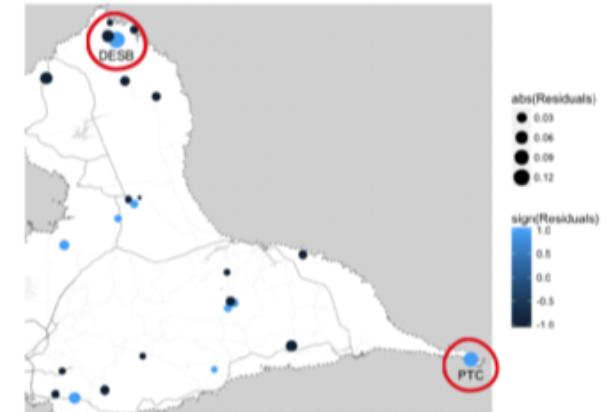

```
qmplot ( Longitude, Latitude, data = df,
         color = sign( Residuals ),
         size = abs( Residuals )) +

geom_text( ... )
```

Map Types from Internet

?ggmap :: get_map
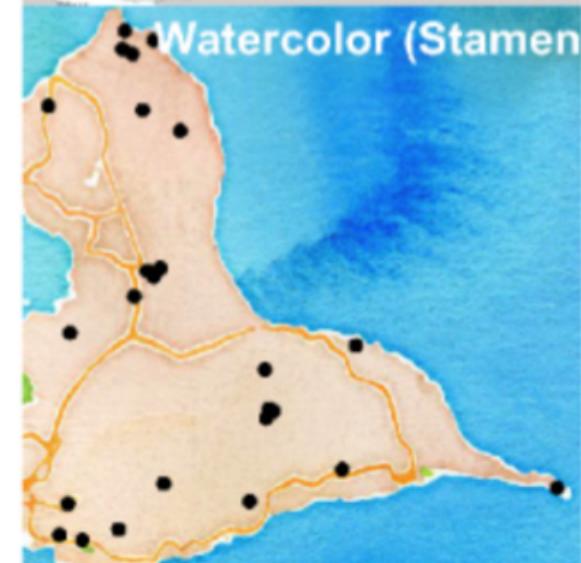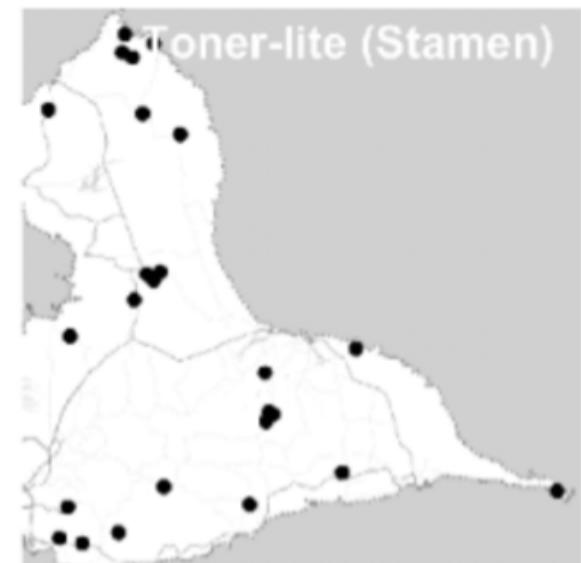
Metapopulation Genetics

Helene Wagner, University of Toronto

**Goal: Test genetic effects of metapopulation dynamics**

- Quantify spatial genetic structure
- Test effects of population size and connectivity
- Test effect of recent extinction events

Methodological Challenges

1. 'EcoGenetics': multi-source data

2. AMOVA with R

3. Regression analysis in R

4. Video 2: R graphics

- Base R graphics

- Grammar of graphics: 'ggplot2'

- Plotting residuals in space

See also Worked Example "Week 0: R Graphics"

D. depressissimum

Photos: J-P Pointier