Zelig: Tools to expand the reach of R

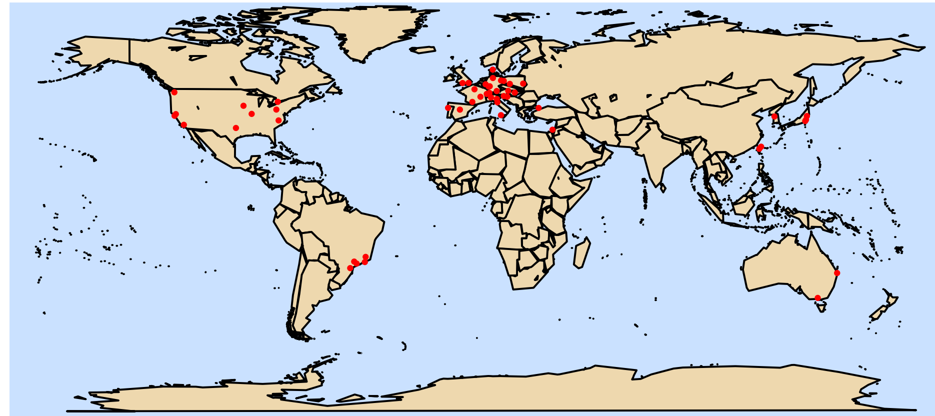
Koskue Imai[1] Gary King[2] Olivia Lau[2]

[1]Department of Politics
Princeton University

[2]Department of Government
Harvard University

Vienna, Austria
June 16, 2006

The current R world



What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
- For researchers, students, and instructors

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
- For researchers, students, and instructors

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands $\longrightarrow$ Simplified UI
- For researchers, students, and instructors
  who don't use R:
  An automatically-generated GUI

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands $\longrightarrow$ Simplified UI
- For researchers, students, and instructors
  who don't use R:
  An automatically-generated GUI

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands $\longrightarrow$ Simplified UI
- For researchers, students, and instructors
  who don't use R:
  An automatically-generated GUI

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands $\longrightarrow$ Simplified UI
- For researchers, students, and instructors
  who don't use R:
  An automatically-generated GUI

What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands → Simplified UI
- For researchers, students, and instructors *who don't use R:*
  An automatically-generated GUI

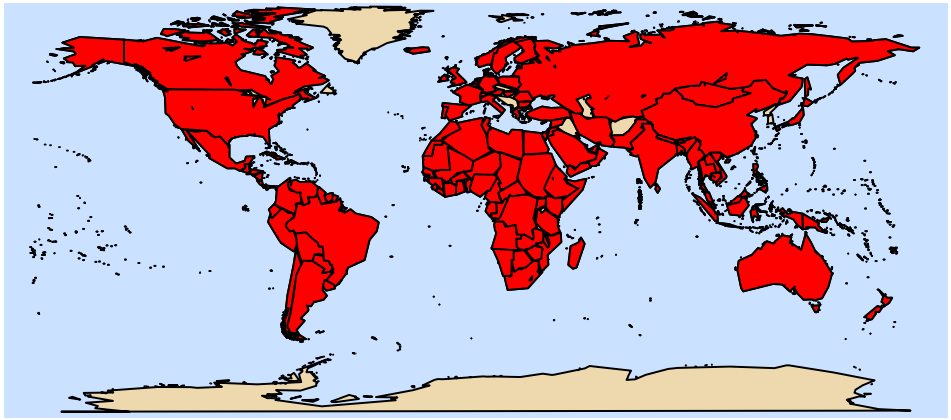
What does Zelig do?

- For developers:
  Makes it easier to translate user-inputs
- For researchers, students, and instructors:
  - Computes quantities of substantive interest for every model

$$E(Y) \quad Y|X \quad E(Y|X_1) - E(Y|X_0) \quad \ldots$$

  - 3 commands → Simplified UI
- For researchers, students, and instructors *who don't use R:*
  An automatically-generated GUI

Goal: R world (with the help of Zelig)



Users only need to know 3 commands

```
z <- zelig(vote~race+educate,      Select vars
     data=turnout,                 Select data set
     model="probit")               Select model

x <- setx(z.out, educate=12)       Select QIs

s <- sim(z, x=x)                   Calculate QIs
```

Users only need to know 3 commands

```
z <- zelig(vote~race+educate,
       data=turnout,
       model="probit")
```
Select vars
Select data set
Select model

```
x <- setx(z.out, educate=12)
```
Select QIs

```
s <- sim(z, x=x)
```
Calculate QIs

Users only need to know 3 commands

```
z <- zelig(vote~race+educate,
       data=turnout,
       model="probit")
```
Select vars
Select data set
Select model

```
x <- setx(z.out, educate=12)
```
Select QIs

```
s <- sim(z, x=x)
```
Calculate QIs

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Users only need to know 3 commands

```
z <- zelig(vote~race+educate,
       data=turnout,
       model="probit")
```
Select vars
Select data set
Select model

```
x <- setx(z.out, educate=12)
```
Select QIs

```
s <- sim(z, x=x)
```
Calculate QIs

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Users only need to know 3 commands

```
z <- zelig(vote~race+educate,
       data=turnout,
       model="probit")
```
Select vars
Select data set
Select model

```
x <- setx(z.out, educate=12)
```
Select QIs

```
s <- sim(z, x=x)
```
Calculate QIs

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Built-in Zelig functionality

- Handle sets of multiply-imputed data frames
- Bootstrap quantities of interest
- Stratify data
- Conditional prediction

Built-in Zelig functionality

- Handle sets of multiply-imputed data frames
- Bootstrap quantities of interest
- Stratify data
- Conditional prediction

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Built-in Zelig functionality

- Handle sets of multiply-imputed data frames
- Bootstrap quantities of interest
- Stratify data
- Conditional prediction

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Built-in Zelig functionality

- Handle sets of multiply-imputed data frames
- Bootstrap quantities of interest
- Stratify data
- Conditional prediction

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

Built-in Zelig functionality

- Handle sets of multiply-imputed data frames
- Bootstrap quantities of interest
- Stratify data
- Conditional prediction

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

No need to change existing packages

Developers only need to add a few functions

Estimate `zelig()`

- (1) `zelig2mymodel()`
- (2) `mymodel()`

Interpret `sim()`

- (3) `param.myclass()`
- (4) `qi.myclass()`

No need to change existing packages

Developers only need to add a few functions

Estimate `zelig()`

- (1) `zelig2mymodel()`
- (2) `mymodel()`

Interpret `sim()`

- (3) `param.myclass()`
- (4) `qi.myclass()`

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

R framework for interpreting user-inputs

- User specifies a `formula`
- `model.frame`
- `model.matrix`
- `model.extract` (or `model.response`)

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

R framework for interpreting user-inputs

- User specifies a `formula`
- `model.frame`
- `model.matrix`
- `model.extract` (or `model.response`)

R framework for interpreting user-inputs

- User specifies a `formula`
- `model.frame`
- `model.matrix`
- `model.extract` (or `model.response`)

R framework for interpreting user-inputs

- User specifies a `formula`
- `model.frame`
- `model.matrix`
- `model.extract` (or `model.response`)

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  → different and diverse UIs
- Challenging for programmers
  → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  → different and diverse UIs
- Challenging for programmers
  → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  → different and diverse UIs
- Challenging for programmers
  → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  → different and diverse UIs
- Challenging for programmers
  → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  → different and diverse UIs
- Challenging for programmers
  → confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Koskue Imai, Gary King, Olivia Lau Zelig: Everyone's Statistical Software

But for more than one equation?

- Hard to parse parameter vector into appropriate subsets in multi-eqn models
- Each package has different hacks
  $\longrightarrow$ different and diverse UIs
- Challenging for programmers
  $\longrightarrow$ confusing for users

Zelig provides tools to extend the R single-equation format to an intuitive multiple-equation UI and API

Represent arbitrarily complicated models

Use a list of equations:

```
f <- list(mu1 = y1 ~ x1 + x2 + x3,
          mu2 = y2 ~ x1 + x4 + x5)

f <- list(mu1 = y1 ~ x1 + tag(x2, beta2),
          mu2 = y2 ~ x3 + tag(x4, beta2),
          rho = ~ z1 - 1)

f <- list(cbind(y1, y2) ~ x1 + x2)
```


Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`

Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`

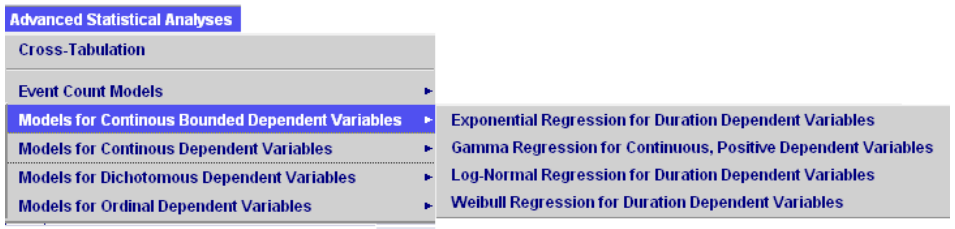
Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`

Developer tools for lists of eqns

- `parse.formula(formula, model)`
- `model.frame.multiple(formula, data)`
- `model.matrix.multiple(formula, data, eqn, shape)`
- `parse.par(par, terms, eqn, shape)`

The Virtual Data Center GUI for Zelig



Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI "dynamically" extends to include new models!
- Developer doesn't need to write a graphical module

Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI "dynamically" extends to include new models!
- Developer doesn't need to write a graphical module

Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI "dynamically" extends to include new models!
- Developer doesn't need to write a graphical module

Model definitions → dynamic GUI

- Developer writes a self-contained function to describe new model
- Every time the VDC is compiled, GUI "dynamically" extends to include new models!
- Developer doesn't need to write a graphical module

Everyone should use R...
...and can with Zelig!

- Partnership with the US Census Bureau (DataFerrett)
- Partnership with the Broad Institute (GenePattern)
- Your suggestions!

Everyone should use R. . .
. . . and can with Zelig!

- Partnership with the US Census Bureau (DataFerrett)
- Partnership with the Broad Institute (GenePattern)
- Your suggestions!

Everyone should use R. . .
. . . and can with Zelig!

- Partnership with the US Census Bureau (DataFerrett)
- Partnership with the Broad Institute (GenePattern)
- Your suggestions!

Visit Zelig on the web at

http://gking.harvard.edu/zelig/