# Using the package “simple features” (`sf`) for sensitivity analysis


Maikol Solís

@maikol_solis

maikol-solis

joint work with:

Alberto Hernández & Ronald Zúñiga

UNIVERSIDAD DE COSTA RICA

@UniversidadCR

CIMPA Centro de Investigación en Matemática Pura y Aplicada

@CIMPAUCR

ESCUELA DE MATEMÁTICA UNIVERSIDAD DE COSTA RICA

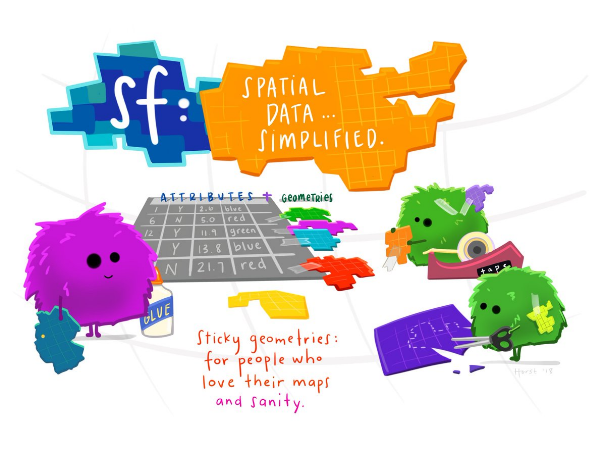
@EscuelaMateUCR


Toulouse R! 2019

Toulouse, France

July 10th, 2019

# `sf` is a great package to work with maps



# Is it possible to use `sf` to solve "statistical" problems not related with maps?

# Statistical Motivation

Assume that $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ produces the output $Y \in \mathbb{R}$ linked by the model

$$Y = \psi(X_1, \ldots, X_p).$$

**The function $\psi$:**

- Could be known or unknown,
- Generally is a complex function.

**Some questions:**

- Is it possible to rank the input variables by its importance?
- Can we measure this importance?

The answer is: Yes!

There are plenty of classical statistical tools to solve this

# What we did new?

We exploited the geometric structure of the data to set the importance of the inputs

# How?

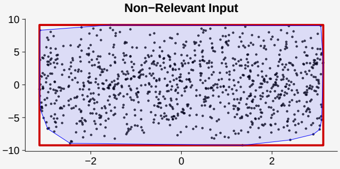
# The geometric correlation

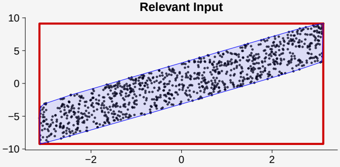
If $X_i$ and $Y$ are geometrically uncorrelated then:

$$\boxed{\text{Box Area}} \approx \boxed{\text{Object Area}}.$$

Otherwise,

$$\boxed{\text{Box Area}} \not\approx \boxed{\text{Object Area}}.$$



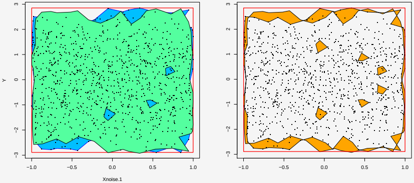

$$\rho_i^{\text{Geom}} = 1 - \frac{\text{Object Area for variable i}}{\text{Box Area for variable i}}.$$

# The geometric sensitivity index

If $X_i$ is geometrical irrelevant to $Y$

Object Area $\approx$ Sym. Reflection

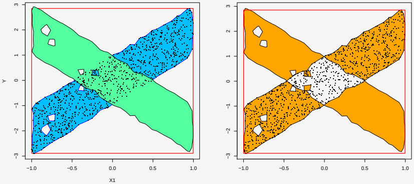
Sym. Difference $\approx 0$




Otherwise,

Object Area $\not\approx$ Sym. Reflection

Sym. Difference $\approx 2 \times$ Object Area




$$S_i^{\text{Geom}} = \frac{\text{Symmetric Difference Area for variable i}}{2 \times \text{Object Area for variable i}}.$$

# Implementation

(The cool stuff!)

# The package `sf`

We use this package to handle two key operations for (Multi)-Polygons:

- Affine transformations.
- Area estimation.

# One example

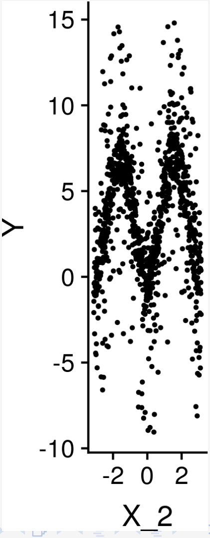
Assume the model

$$Y = \sin X_1 + 7\ \sin^2 X_2 + 0.1\ X_3^4 \sin X_1$$

where $X_i \sim \mathrm{Uniform}(-\pi, \pi)$ for $i = 1, 2, 3$.

**Example**

The points $(X_2, Y)$ look like this:

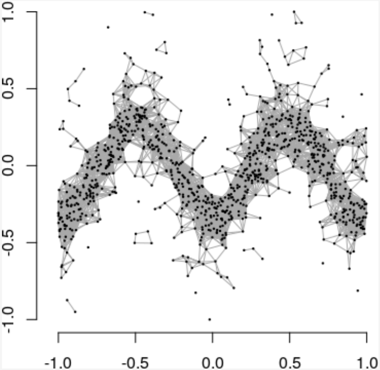


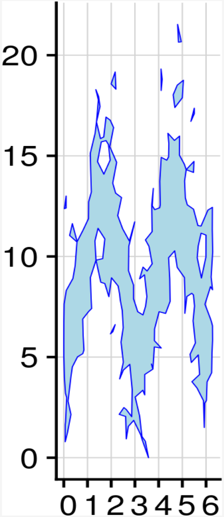
# Estimating the Object

(Embedding manifold)

# Estimating the embedding manifold

Based on the work of Zomorodian we build a Vietoris-Rips Complex of the data.




A. Zomorodian. 2010. Fast construction of the Vietoris-Rips complex. *Computers & Graphics* 34 (3): 263–271

If `l` is the list of triangles in the Vietoris-Rips complex

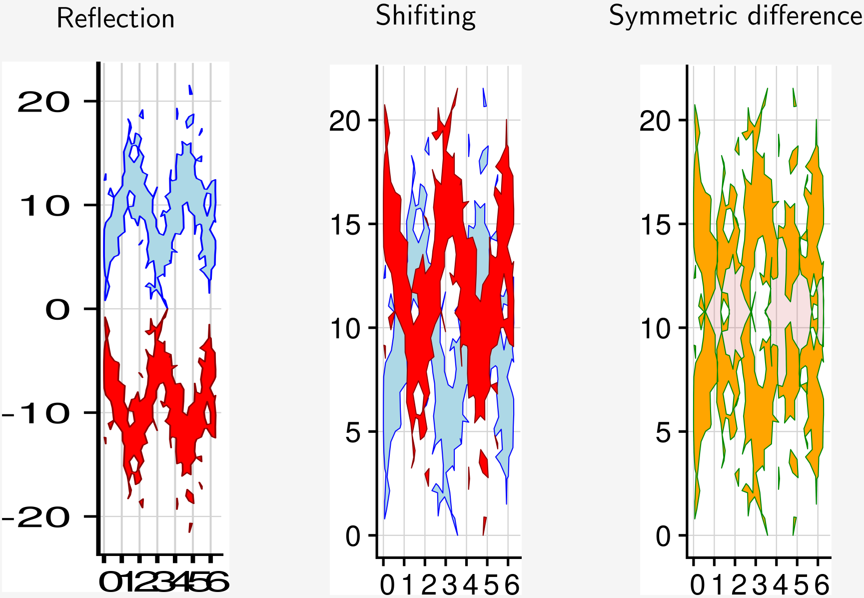
```
VR <- sf::st_multipolygon(l)
VR_union <- sf::st_union(VR)
```

We center the estimated manifold into the origin,

```
VR_origin_coords <- sf::st_coordinates(VR_union)

VR_union <- VR_union -
                 c(min(VR_origin_coords[,"X"]),
                 min(VR_origin_coords[,"Y"]))
```

# Estimating the Symmetric Reflection


Reflection
20
10
0
-10
-20
0123456
Shifiting
20
15
10
5
0
0 1 2 3 4 5 6
Symmetric difference
20
15
10
5
0
0 1 2 3 4 5 6

## Reflection

Multiplying the manifold by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

```
reflection <-  matrix(c(1, 0, 0, -1), 2, 2)
VR_reflection <- (VR_union * reflection)
```

## Shifting

```
VR_coordinates <- sf::st_coordinates(VR_union)
VR_reflection <- VR_reflection +
                        c(0, 2 * mean(VR_coordinates[, "Y"]))
```

## Sym. difference

```
VR_symmetric_difference <- sf::st_sym_difference(VR_union,
                                                VR_reflection)
```

**Example**

All the code is available in the package `topsa` (Topological Sensitivity Analysis)

`www.github.com/maikol-solis/topsa`

# Some numerical examples

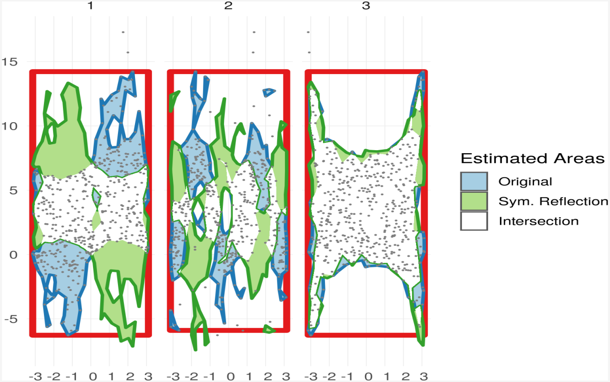
# The Ishigami model

$$Y = \sin X_1 + 7\ \sin^2 X_2 + 0.1\ X_3^4 \sin X_1$$

where $X_i \sim \mathrm{Uniform}(-\pi, \pi)$ for $i = 1, 2, 3$.

```
ishigami <- topsa::topsa(Ydat = Y,
                         Xdat = X,
                         method = "VR")
topsa::plot(ishigami)
topsa::print(ishigami)
```

# The Ishigami model




| | Max. Radius | Manifold | Box | Correlation | Sensitivity |
|---|---|---|---|---|---|
| $X_1$ | 0.09 | 55.72 | 128.01 | 56% | 46% |
| $X_2$ | 0.08 | 39.81 | 126.28 | 68% | 57% |
| $X_3$ | 0.09 | 61.25 | 128.47 | 52% | 11% |

**True sensitivity (Sobol indices):** $X_1 = 31\%$ $X_2 = 44\%$ $X_3 = 0\%$

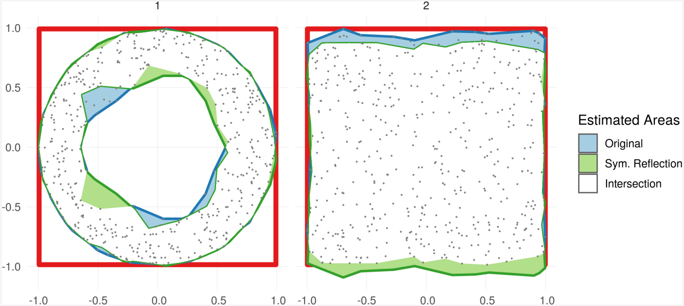
# The Donut model

$$\begin{cases} X_1 = r\cos(\theta) \\ Y = r\sin(\theta) \end{cases}$$

with $r \sim \mathrm{Unif}(0.5, 1)$ and $\theta \sim \mathrm{Unif}(0, 2\pi)$ random. The variable $X_2$ is pure noise.

```
donut <- topsa::topsa(Ydat = Y,
                      Xdat = X,
                      method = "VR")
topsa::plot(donut)
topsa::print(donut)
```

# The Donut model



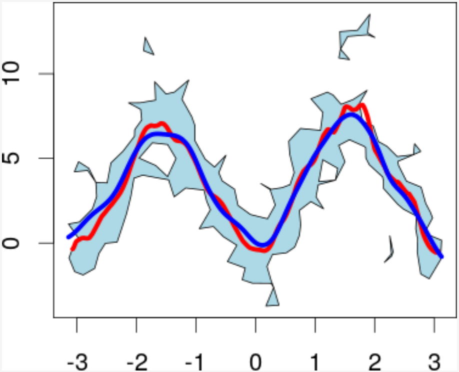

| | Max. Radius | Manifold | Box | Correlation | Sensitivity |
|---|---|---|---|---|---|
| $X_1$ | 0.16 | 1.91 | 3.92 | 51% | 8% |
| $X_2$ | 0.20 | 3.73 | 3.94 | 5% | 5% |

**True sensitivity (Sobol indices):**

$$X_1 = 0\% \qquad X_2 = 0\%$$

# What is next?

# Fit a topological regression curve




Non-parametric kernel regression    Topological curve regression

# Optimize the construction of the complex

(persistence homology)

# Send it to CRAN
(after iron out some vicious bugs...)

Maikol Solís
@maikol_solis
maikol-solis

joint work with:

Alberto Hernández & Ronald Zúñiga

## Preprints

Hernández, Alberto, Maikol Solís, and Ronald Zúñiga. "Geometrical correlation indices using homological constructions on manifolds". Submitted to be published. (2018).

Hernández, Alberto, Maikol Solís, and Ronald Zúñiga. "Sensitivity indices on homological constructions through symmetric reflections". In preparation. (2019).