

Teaching Data Science in Adverse Circumstances

Posit Cloud and Quarto to the Rescue

Aleksander Dietrichson, PhD



Argentina

- Population: ~ 45 M
- GDP/Capita: ~13K
- Poverty rate: ~40%
- Inflation: 113.4%

Buenos Aires



UNSAM



Furthermore...

- $< 42\%$ of students have access to Laptops/Computers
- $> 70\%$ Work while studying
- «Escuela de Humanidades» – Liberal Arts College

Challenges

- Lack of Resources
- Language Barriers
- Arithmophobia
- Technophobia

Why Care?

- Lack of resources should not hinder entrance
- Nor linguistics
- Journalists should be statistically literate (critical)
- PR people too
- English majors need jobs too
- Diversity enriches our community

- Free version available
- Stable hardware and config
- Runs on anything...

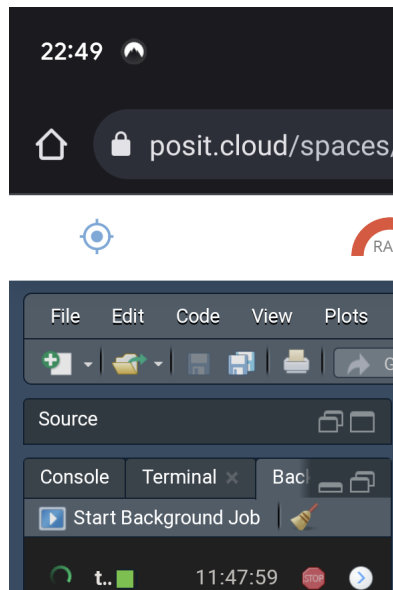
Solves the resource issue



Posit.cloud

- Free version available
- Stable hardware and config
- Runs on anything...

Solves the resource issue



Language Barriers

In R itself:

Language Barriers

In R itself:

```
mis_datos <-  
  coches |>  
  elegir(cil, disp, fc) |>  
  agrupar_por(cil)|>  
  resumir(  
    disp = promedio(disp),  
    fc = promedio(fc)  
  )  
  
imprimir(mis_datos$disp)
```

Language Barriers

In R itself:

```
mis_datos <-  
  coches |>  
  eligir(cil, disp, fc) |>  
  agrupar_por(cil)|>  
  resumir(  
    disp = promedio(disp),  
    fc = promedio(fc)  
  )  
  
imprimir(mis_datos$disp)
```

```
my_data <-  
  mtcars |>  
  select(cyl, disp, hp) |>  
  group_by(cyl)|>  
  summarise(  
    disp = mean(disp),  
    hp = mean(hp)  
  )  
  
print(my_data$disp)
```

Language Barriers

In R itself:

```
mis_datos <-  
  coches |>  
  eligir(cil, disp, fc) |>  
  agrupar_por(cil)|>  
  resumir(  
    disp = promedio(disp),  
    fc = promedio(fc)  
  )
```

```
imprimir(mis_datos$disp)
```

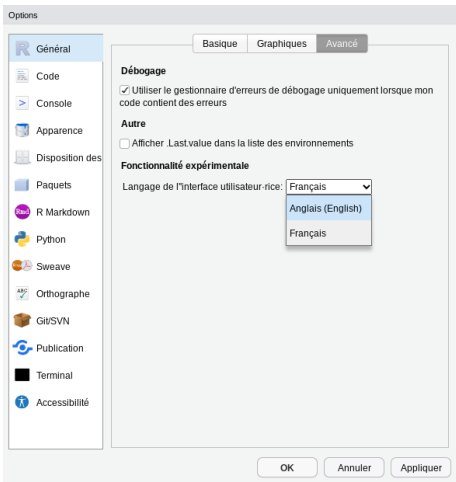
```
my_data <-  
  mtcars |>  
  select(cyl, disp, hp) |>  
  group_by(cyl)|>  
  summarise(  
    disp = mean(disp),  
    hp = mean(hp)  
  )
```

```
print(my_data$disp)
```

- Solution impractical / counter-productive
- Steeper learning curve

Language Barriers

Interface:



- As of September 2023
- Experimental feature
- Two languages available
 - English
 - French

Language Barriers

Teaching materials:

Métodos Cuantitativos

Prefacio

- 1 Conceptos fundamentales
- 2 Distribuciones de frecuencias
- 3 Centralización y dispersión
- 4 La distribución normal
- 5 Estimación de parámetros
- 6 Diseño de proyectos y test de hipótesis
- 7 Pruebas paramétricas
- 8 Pruebas no paramétricas
- 9 Prueba de χ^2
- 10 Correlación

Referencias

Apéndices

- A Distribución t
- B Valores críticos del test de

Métodos Cuantitativos

Materiales de cátedra y de consulta

AUTOR/A

Aleksander Dietrichson, PhD

FECHA DE PUBLICACIÓN

15 de enero de 2023

Prefacio

Este texto ha sido editado en respuesta a la aparente falta de un libro de texto introductorio al análisis cuantitativo y estadísticas accesible y moderno en castellano. Si bien fue concebido como material de cátedra para *Metodologías cuantitativas* materia que dicta el autor en la Escuela de Humanidades de la Universidad Nacional San Martín, se adaptará fácilmente a cursos introductorios de estadísticas en general.

Segunda edición

En la segunda edición (2021) se corrigió algunos errores ortográficos y de estilo. Optamos por actualizar los ejemplos para incorporar los paquetes del «tidyverse» ya que hemos observado que su uso y adaptación atenúa la curva de aprendizaje para quienes usan R por primera vez o con escasos conocimientos previos.

Tabla de contenidos

[Prefacio](#)

- Segunda edición
- Tercera edición
- Estructura del libro
- R y Rstudio
- Edición
- Licencia
- Agradecimientos

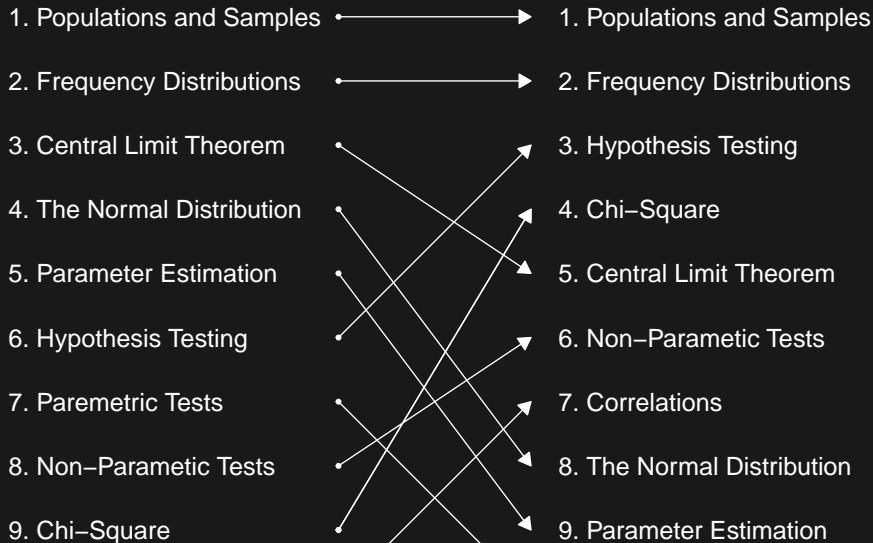
Arithmophobia

- Curriculum reorganization
- Use of relevant data
- Pedagogical devices

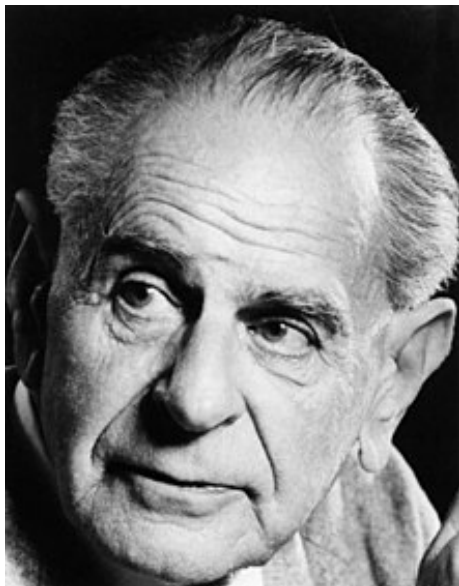
Curriculum

1. Populations and Samples
2. Frequency Distributions
3. Central Limit Theorem
4. The Normal Distribution
5. Parameter Estimation
6. Hypothesis Testing
7. Parametric Tests
8. Non-Parametric Tests
9. Chi-Square

Curriculum



Hypothesis Testing



- The Scientific Method
- Falsifiability
- Hypothesis Testing
- Paradigm Shift

Chi Square

- Requires four numbers
- Nominal data

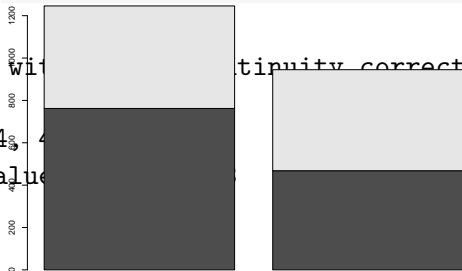
- Visualize with a line
- Base R

```
rbind(c(762, 468), c(484, 477))
chisq.test()
```

```
rbind(c(762, 468), c(484, 477))
barplot()
```

Pearson's Chi-squared test with continuity correction

```
data: rbind(c(762, 468), c(484, 477))
X-squared = 29.06, df = 1, p-value = 1.11e-06
```



Relevant Data

```
indec_data |>
  group_by(CH04) |> # CH04 = Gender: 1 = Male, 2 = Female
  summarize(
    Median = median(ITF), # Total Family Income
    Mean = mean(ITF),
    SD = sd(ITF)
  )
```

...

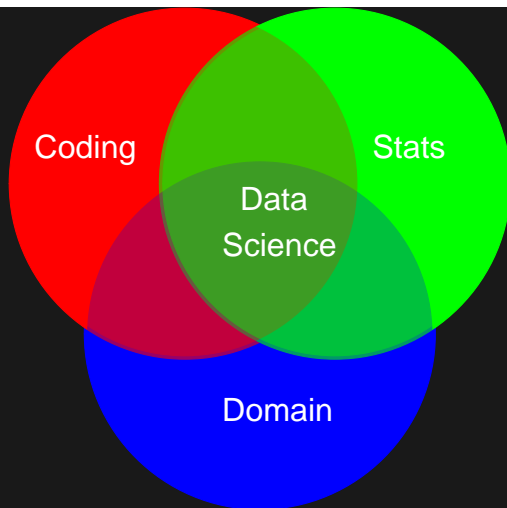
Stat	1	2
Median	145000	136500
Mean	175743	166718
SD	230829	184870

Source: Instituto Nacional de Estadística y Censos (2023)

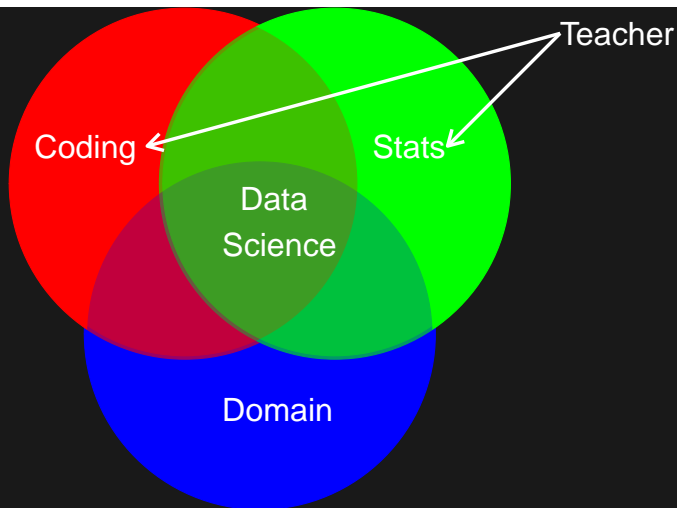
Pedagogical Devices

- Live coding
- Schadenfreude

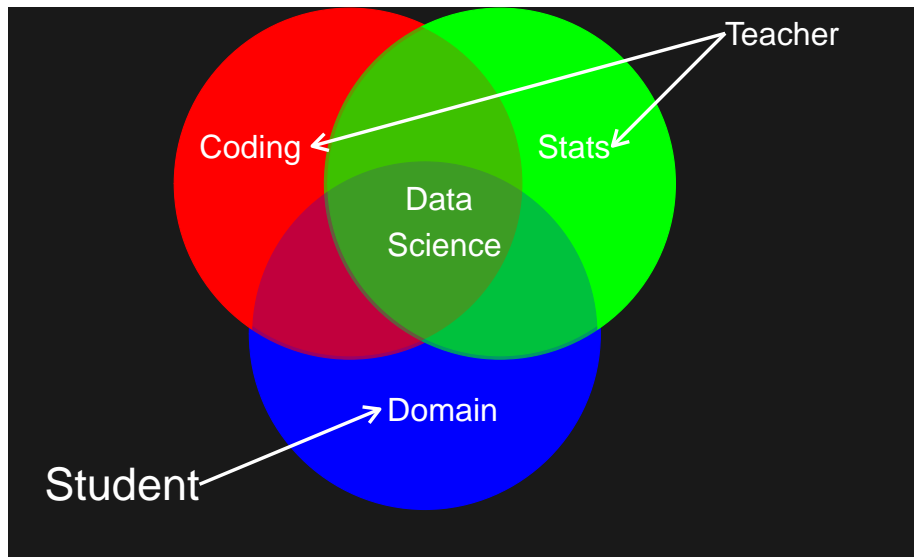
What We Teach



What We Teach



Student is the Expert



Student is the Expert

Matías Salto:

- Billboards in the BA subway
- Warm/Cold colors ~ product
- Collects data on cellphone

```
image |>
```

```
as.raster() |>
```

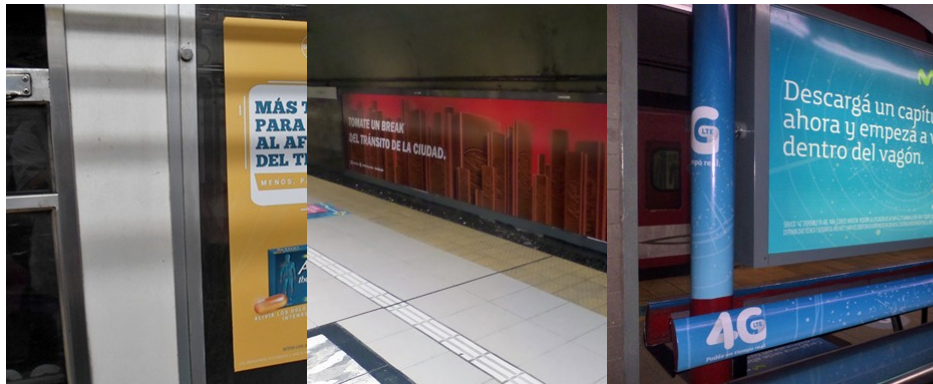
```
as.vector() |>
```

```
head(10)
```

```
[1] "#938e8bff" "#95908dff" "#  
[7] "#938f8cff" "#938f8cff" "#
```



Student is the Expert



$$\text{logit}(p) = \beta_0 + \beta_1 \text{red} + \beta_2 \text{green} + \beta_3 \text{blue}$$

Student is the Expert

Camila Ramírez:

- Animated Shorts
- Visual language
- Color as a semiotic resource



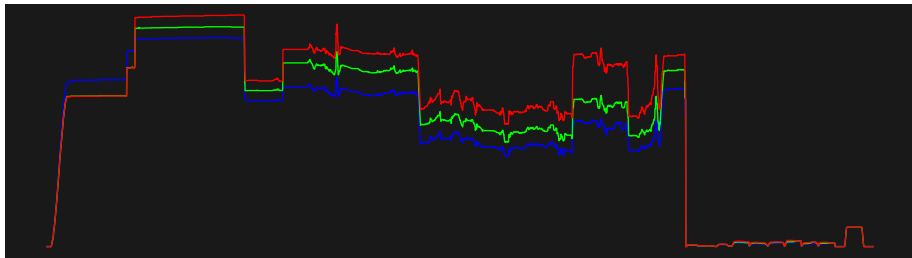
Student is the Expert

Camila Ramírez:

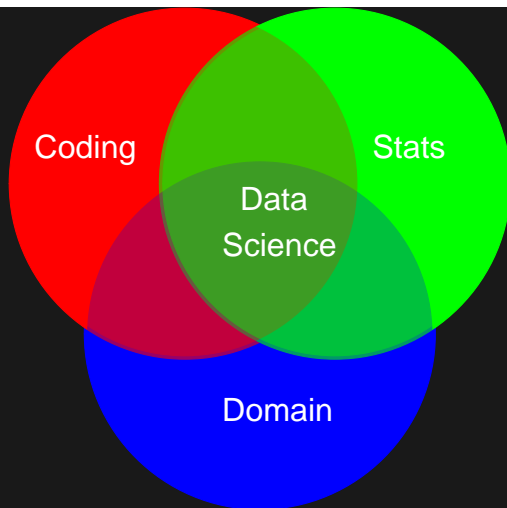
- Animated Shorts
- Visual language
- Color as a semiotic resource



Student is the Expert



Paradigm Shift



Paradigm Shift

Coding

Stats

Domain

Conclusions

- Be relevant, be real, be there
 - Curriculum
 - Teaching materials
 - Data
 - Pedagogy
- Keep supporting those who teach Data Science

