

7 Collaborate and communicate

(studiare da soli)

Questa è la parte più importante perché l'esposizione è la valutazione del lavoro. Comunicare in modo efficace ed accattivante è un'arte che si può imparare. La prima regola è l'eticità. Se così non fatto chi ascolta si sente fregato.

Parafrasando Ipocrate:

- Se non si è riusciti a far qualcosa durante il progetto o si dice oppure non si cela questa cosa
- Quando si fanno delle analisi commissionate bisogna rispettare la privacy, quindi non esporre dati sensibili.
- Ricordarsi che ci sono obblighi in qualità di essere umano e quindi anche in questo caso avere una forma di rispetto.

6.6 Riproducibilità dell'analisi:

Un'altra cosa che ha a che fare con l'etica e la riproducibilità delle analisi. Quando si fa un'analisi dei dati si registra ogni passo anche i fallimenti. In questo il formato markdown è perfetto. È molto importante documentare l'analisi tenendo traccia, questo perché altri potrebbero voler riutilizzare l'analisi. Fondamentalmente si mette a disposizione il codice e il DSet. Per esempio, Excel non tiene traccia del percorso e quindi sarà difficile condividere il percorso fatto.

6.8 sorprendentemente (publication bias) pubblicare poco non è etico.

Alle volte non pubblicare qualcosa è poco etico. Questo ha a che fare con il "publication bias", cioè fenomeno per il quale vengono riportati solo i successi e non i fallimenti. L'effetto è che apparentemente si hanno solo risultati positivi, anche se sono solo il 5% delle prove fatte. Quindi potremmo avere un'idea sbagliata perché il 95 % degli esperimenti sul medesimo farmaco ad esempio sono falliti. Quindi la prima regola è la correttezza. Un'altra cosa è collaborare, quindi mettere apertamente i dati a disposizione e magari mettere anche il codice a disposizione. Un modo per farlo è Git e Github. Questo è lo strumento per pubblicare il progetto. Quello che si consiglia è di guardare il corso su DataCamp che riassume tutta la prossima parte.

Fare su datacamp working with R studio capitolo 2: git e github e parte 3 (x reporting: cioè r markdown) x esame

7.4 Interattività

Shiny è un modo per rendere interattiva le presentazioni. In R è possibile gestire l'interattività con Shiny e **HTML widget**. Questi ultimi sono dei componenti che possiamo utilizzare nel vostro markdown e che rendono le nostre visualizzazioni interattive. Queste componenti vengono elaborati a livello di client. Vengono convertiti in un codice javascript che poi funziona con un qualsiasi browser moderno. Per vedere esempi di vada sul link http://www.htmlwidgets.org/showcase_leaflet.html

L'ultima (e terza) componente che si possono usare sono i **dashboard** che sono dei cruscotti. Per degli esempi: <https://rmarkdown.rstudio.com/flexdashboard/examples.html>

Questo metodo ha uno svantaggio, perché si rendono interattive solo le visualizzazioni che hanno un grafico. Se vogliamo creare noi una nostra visualizzazione dobbiamo usare Shiny. In questo caso costruiamo noi l'interazione, quindi sia il widget che i componenti reattivi. La computazione avviene sul server. Per vedere degli esempi si vada qui: <http://shiny.rstudio.com/gallery/>

Qua non c'è nulla di javascript, ma la computazione avviene sul server. Lo svantaggio quindi è quello di dover aver un server. Noi possiamo far girare con R Studio, esempio:

7.4.1 shiny

```
library(shiny)

# Define UI for the application
# Fluid pages scale their components in realtime to fill all available browser width.
ui = fluidPage(
```

```

# Application title
titlePanel("Hello Shiny"),

# Create a layout with a sidebar and main area. The sidebar is displayed with a distinct
background color and
# typically contains input controls. The main area occupies 2/3 of the horizontal width and
typically contains outputs.
sidebarLayout(
  # sidebar panel
  sidebarPanel("Sidebar panel"),

  # main panel
  mainPanel("Main panel")
)

# Define server logic
server = function(input, output) {}

# Run the application
shinyApp(ui = ui, server = server)

```

Un altro esempio

```

library(shiny)

# Define UI with a basic widgets
ui = fluidPage(
  titlePanel("Basic widgets"),

  fluidRow(

    column(3,
      h3("Buttons"),
      actionButton("action", "Action"),
      br(),
      br(),
      submitButton("Submit")),

    column(3,
      h3("Single checkbox"),
      checkboxInput("checkbox", "Choice A", value = TRUE)),

    column(3,
      checkboxGroupInput("checkGroup",
        h3("Checkbox group"),
        choices = list("Choice 1" = 1,
                      "Choice 2" = 2,
                      "Choice 3" = 3),
        selected = 1)),

    column(3,
      dateInput("date",
        h3("Date input"),
        value = "2014-01-01"))
  ),

  fluidRow(

    column(3,
      dateRangeInput("dates", h3("Date range"))),

    column(3,
      fileInput("file", h3("File input"))),

    column(3,
      h3("Help text"),
      helpText("Note: help text isn't a true widget,",
        "but it provides an e
        asy way to add text to",
        "accompany other widgets.")),

    column(3,
      numericInput("num",
        h3("Numeric input"),
        value = 1))
  )

```

```

),

fluidRow(

  column(3,
    radioButtons("radio", h3("Radio buttons"),
      choices = list("Choice 1" = 1, "Choice 2" = 2,
        "Choice 3" = 3), selected = 1)),

  column(3,
    selectInput("select", h3("Select box"),
      choices = list("Choice 1" = 1, "Choice 2" = 2,
        "Choice 3" = 3), selected = 1)),

  column(3,
    sliderInput("slider1", h3("Sliders"),
      min = 0, max = 100, value = 50),
    sliderInput("slider2", "",
      min = 0, max = 100, value = c(25, 75))
  ),

  column(3,
    textInput("text", h3("Text input"),
      value = "Enter text...")
  )
)

# Define server logic
server = function(input, output) {}

# Run the app
shinyApp(ui = ui, server = server)

```

7.5 infografiche

Se voglio uscire dagli standard: 4 autori infografici Edward Tufte, Moritz Stefaner, Stefanie Posavec, Giorgia Lupi

7.5.1 Edward Tufte

Edward Tufte è uno statista e artista e professore emerito di scienze politiche, statistiche e informatica presso la Yale University. Ha scritto, progettato e auto-pubblicato 4 libri classici sulla visualizzazione dei dati. È noto per i suoi scritti sulla progettazione delle informazioni e come pioniere nel campo della visualizzazione dei dati.

Ha coniato la parola *chartjunk* per riferirsi a elementi inutili, non informativi o oscuranti le informazioni di visualizzazioni di informazioni quantitative.

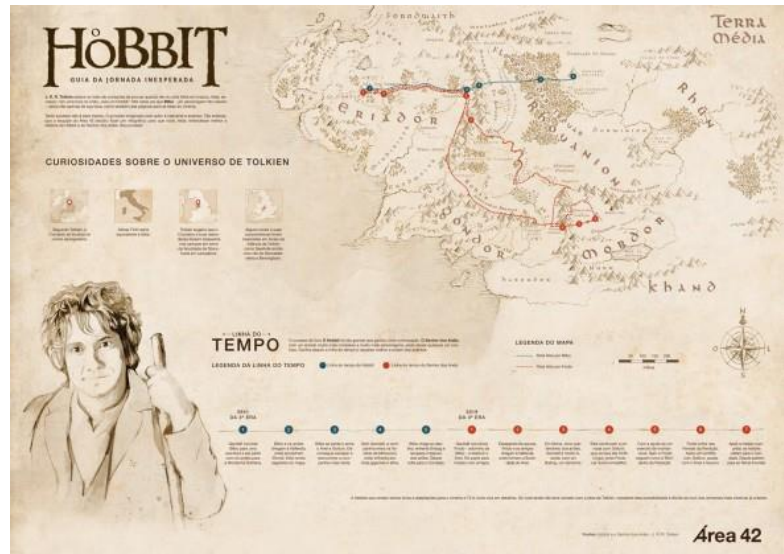
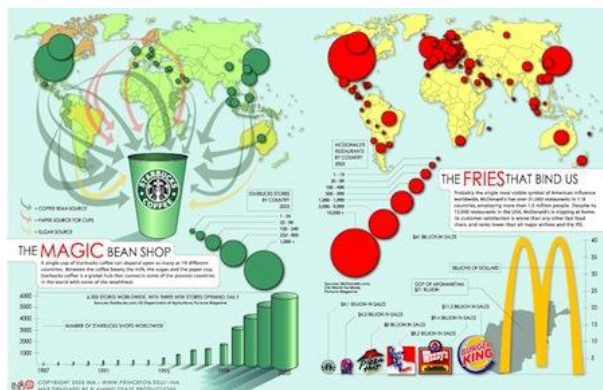
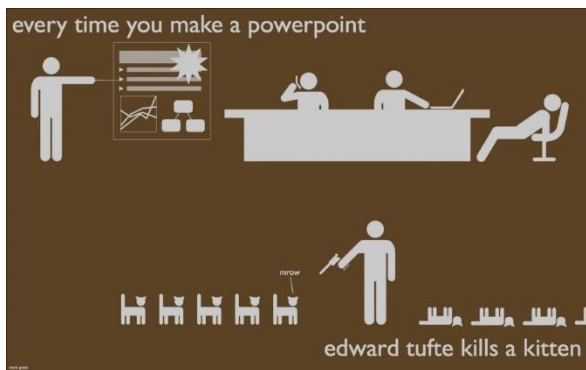
Sparkline: anche se si dice che Tufte abbia inventato le *sparkline*, in realtà inventò solo il nome e lo rese popolare come tecnica. Le sparkline sono un modo condensato per presentare tendenze e variazioni, associate a una misurazione come la temperatura media o l'attività di borsa, spesso incorporate direttamente nel testo; per esempio: l'indice Dow Jones per il 7 febbraio 2006 sparkline che illustra le fluttuazioni nell'indice di Dow Jones il 7 febbraio 2006



Tufte incoraggia l'uso di illustrazioni ricche di dati che presentano tutti i dati disponibili. Quando tali illustrazioni vengono esaminate attentamente, ogni punto di dati ha un valore, ma quando vengono guardati più in generale, si possono osservare solo tendenze e modelli. Tufte suggerisce che queste letture macro / micro siano presentate nello spazio di un occhio, nel formato ad alta risoluzione della pagina stampata e al ritmo lento del tempo libero dello spettatore.



Esempi di infografiche:



7.5.2 Moritz Stefaner

Vivo e respiro la visualizzazione dei dati - come designer indipendente, consulente e ricercatore. Aiuto organizzazioni, ricercatori e aziende a trovare verità e bellezza in dati rilevanti e significativi. Ho lavorato con clienti come FIFA, Skype e DPA e ho rapporti di consulenza a lungo termine con l'OCSE, la Deutsche Bahn e la Max Planck Research Society.

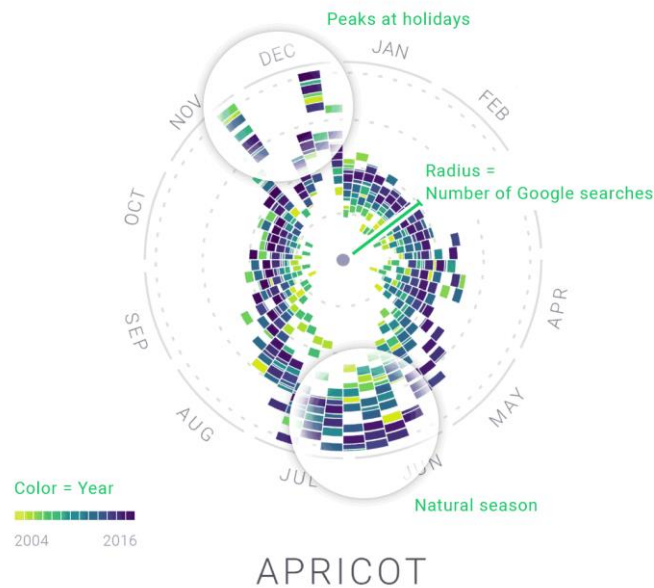
Una delle opere più note di Stefaner è la visualizzazione del Better Life Index dell'OCSE. Tra gli altri progetti degni di nota c'è la sua installazione interattiva On Broadway, il suo lavoro per la FIFA, e la sua progettazione del nuovo portale di dati dell'OCSE

Esempio di visualizzazione:

Cosa possiamo imparare sulla cultura del cibo guardando i dati di ricerca di Google? Quando la gente cerca frullati, zuppe o macedonia? *The Rhythm of Food*, un progetto di visualizzazione in collaborazione con Google News Lab, mette in luce le molte sfaccettature della stagionalità alimentare, basate su dodici anni di dati di ricerca di Google.

Al fine di investigare i modelli stagionali nelle ricerche alimentari, abbiamo sviluppato un nuovo tipo di grafico radiale a "anno" per rivelare le tendenze stagionali degli alimenti. Ogni segmento del grafico indica l'interesse di ricerca in una delle settimane degli ultimi 12 anni, con la sua distanza dal centro che mostra l'interesse di ricerca relativo e il colore che indica l'anno. Ciò consente di individuare entrambi i ritmi che si ripetono su base annuale (come la stagione naturale, i picchi durante le vacanze, ...) ma anche le tendenze anno su anno (come l'aumento dell'avocado o il crollo dell'interesse nelle bevande energetiche).





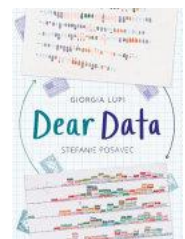
7.5.3 Stefanie Posavec

Nata nel 1981, vive a Londra. È una *information designer* il cui lavoro si concentra su rappresentazioni di dati non tradizionali. È stata coautrice del libro 2016 *Dear Data* con Giorgia Lupi che è il culmine di un progetto di un anno che traduce i dati personali in visualizzazioni disegnate a mano. Le sue opere si trovano nei musei di vari Paesi.

Dear Data è una raccolta di cartoline contenenti dati registrati dalla vita quotidiana dei designer dell'informazione Stefanie Posavec e Giorgia Lupi.



La premessa di questo progetto era testare in prima persona quanto bene si potesse conoscere qualcuno leggendo i propri dati. Infatti le due autrici si incontrarono solo due volte, prima degli scambi di cartoline. Le cartoline venivano scambiate tra loro su base settimanale con i dati rappresentati tramite un disegno su un lato e una chiave dettagliata o una legenda sull'altro. Alcuni degli argomenti includono: Una settimana di pensieri negativi, Una settimana di uno spazio di lavoro e Una settimana di cose nuove.



7.5.4 Giorgia Lupi

Laureata in architettura a Ferrara, ha lavorato con aziende di design in Italia. Vive a NY dal 2012. È una designer dell'informazione italiana e co-fondatrice della società di ricerca e design basata sui dati Accurat, azienda che combina design e dati per creare visualizzazioni e interfacce, che ha avuto clienti come Google, Microsoft, OMS, NYT.



Data humanism: introdotto nel 2017 in suo articolo, la Lupi sostiene un approccio più "unico, contestuale e personale" nei confronti dei dati. Si oppone ad alcuni aspetti convenzionali sui dati - come essere grandi, infallibili, imparziali, descrittivi e predittivi - per difendere altri aspetti che ritiene più umani (dati piccoli, dati imperfetti, dati soggettivi, dati stimolanti, dati fortuiti, ecc.).

Spiegazione esame

Esame scritto 40%

SCRITTO 6 domande, 90 minuti di tempo, esercizi “scrivi in dyplyr questa interrogazione”, quali sono le caratteristiche dei big data, potete utilizzare tutti i cheatsheets e solo quelli

Gli interessa che abbiamo capito la semantica non la sintassi, portare quello che serve per ricordare la sintassi

Progetto 60%

No dataset piccolo, significativo e (*) significativo nelle domande, sceglietene uno che vi interessa, fare domande e dare risposte e inviare una settimana prima della esame, una breve relazione di 15 pg: dataset, obbiettivi ricerca, analisi fatte e risultati, no codice, solo cose scritte sulle istruzioni, il report deve essere in Markdown. Separatamente aggiungere tutto il codice scritto con eventualmente il dataset utilizzato.

Caricare il progetto su github.

Portare il portatile per presentare progetto, ho mezz’ora di tempo al massimo per convincere che ho risposto alle domande. La parte più importante del corso è la presentazione *Collaborate and communicate*, consiglio grafici interattivi.

Presentazione in powerpoint, o altro.