



Interactive data science for biotech

Oana C. Florean

Data Scientist

Nagarro



About me

- Graduated biochemical engineering
- Learned neurochemistry, Python and ML at Stockholm Uni
- Worked in Horizon 2020 research projects

Observed that R is preferred in

- Statistics and academic research- powerful for hypothesis testing, statistical analysis, publishing results
- Biotech and pharma- specialized frameworks for biological data, Bioconductor



R

- Developed in the '90s, inspired by S of the '70s at Bell Labs
- Interpreted and scripting language
- Wide variety of advanced statistical and graphical methods
- CRAN packages: dplyr- data manipulation, tidyr- data organization, caret- ML workflows, ggplot2- complex plots, plotly- interactive plots, tidyverse: suite for data science, shiny: interactive web apps
- 1D (vector, factor, list), 2D (matrix, data frame), and nD (array)
- Quirky syntax: assignment “<-”, pipe “%>%”, data structures access “\$”
- Open-source, active community and plenty of resources





Shiny

- Package to create interactive web apps
- Quickly and easily create dashboards directly from R
- Turn data into visualizations, analyses and products

Features

- Structured- the code is organized in separate UI and server sections
- Flexible- pre-built widgets or customized with HTML, CSS, JavaScript
- Reactive- the output automatically updates when the inputs change
- Accessible- hosted on a server and opened through any browser



Shiny

```
# Load Shiny package
library(shiny)

# Define the UI
ui <- fluidPage(
  titlePanel("Neuro Cells Guide"),
  sidebarLayout(
    sidebarPanel(
      selectInput(inputId = "cellType",
                  label = "Cell Type:",
                  choices = c("Neuron",
                             "Astrocyte",
                             "Microglia"),
                  selected = "Neuron")
    ),
    mainPanel(
      textOutput("infoText")
    )
  )
)

# Define server logic
server <- function(input, output) {

  output$infoText <- renderText{

    descriptions <- list(
      Neuron = "Neurons: Signal transmitters.",
      Astrocyte = "Astrocytes: Neuron supporters.",
      Microglia = "Microglia: CNS immune defenders."
    )

    descriptions[[input$cellType]]
  }
}

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

The screenshot shows a Shiny application window titled "Neuro Cells Guide". On the left, there is a sidebar panel containing a dropdown menu labeled "Cell Type" with "Neuron" selected. To the right of the sidebar, the main panel displays the text "Neurons: Signal transmitters.".



Interactive data science

Combines

- User interfaces- designed for clarity and insights
- Real-time exploration- adjust parameters and see results instantly
- Dynamic visualizations- charts that update with the data
- Collaborative work- multiple users, shared data

Exactly what biotech scientists need

- Complex data and analyses accessible to non-programmers
- Joint research accelerating discovery and innovation



The project

Requirements

- Platform for collaborative research
- Multiple analyses for data exploration
- Prototype in Shiny

Solution

- Comprehensive document app
- Versatile analysis modules

Dev team

- Software architect
- Data scientists- integrating statistics; iterative improvement
- Biostatisticians



Alzheimer's

A neurodegenerative disorder leading to cognitive decline and dementia in older adults.

Biomarkers in cerebrospinal fluid

- Insights about disease progression and effects
- Neurogranin (synaptic plasticity), NfL (axon structure), α -Synuclein (neurotransmitters regulation), GFAP, S100B (astrocytes), IL-6, sTREM2 (inflammation)



Alzheimer's research

<https://www.alzheimersdata.org/>

The Alzheimer's Disease Data Initiative (AD Data Initiative) is a coalition of leading academic, advocacy, government, industry, and philanthropy organizations that recognizes the need for dementia researchers to find easier ways to share unpublished data, analytical tools, and scientific findings. These partners are working together to accelerate progress towards new diagnostics, treatments, and cures in Alzheimer's disease and related dementias (ADRD).

Alzheimer's Disease Data Initiative Coalition

Led by a global coalition of partners, the AD Data Initiative is empowering researchers worldwide by enabling seamless access to multiple data sharing platforms, unlocking important ADRD datasets and fostering research collaboration.



AD Workbench

Powering breakthroughs in Alzheimer's and related dementias research

AD Workbench – a secure, cloud-based data sharing and analytics environment – is the interoperability layer of the AD Data Initiative technical suite and its flagship product offering. Easy to use and available at no cost, AD Workbench empowers researchers around the world to share, access and analyze data across platforms.

AD Workbench Login

Submit Data

Analysis App



The screenshot shows the NTK Analysis App interface. At the top, there is a navigation bar with links to AD Workbench, ntk-datahackathon, FAIR Data Services, AD Connect, a search bar, and a user profile for DANA.FLOREANIN. Below the navigation bar is a blue header bar with links to Home, Activity, Files, Database, Tools, Apps, Upload, and Administer. The main area is titled "NTK Analysis App" and contains a "Document" tab. The toolbar includes buttons for LOAD DATASET, DATA, DICTIONARY, ANALYSIS, LOAD ANALYSIS, USE TEMPLATE, SAVE, SAVE AS TEMPLATE, and MODULE CREATOR. There are also buttons for Paragraph, Bold, Italic, Underline, and various table and list options. A "Display module configuration" checkbox is checked. On the left, a "Table of contents" section has a "show paragraphs" checkbox. The main content area contains the text "Let's conduct an analysis!". On the right, there is a "Document recovery" section with a timestamp: "Recover time is GMT+0: 2024-03-14 16:49:08 GMT" and "2024-03-14 16:31:50 GMT".

EDA module example 1



The screenshot shows the AD Workbench interface with the following details:

- Top Bar:** AD Workbench, ntk-datahackathon, FAIR Data Services, AD Connect, Search bar, Help icon, and User name (DANA.FLOREANIS).
- Header:** Home, Activity, Files, Database, Tools, Apps, Upload, Administer.
- Current Application:** Apps, NTK Analysis App.
- Toolbar:** Document (with a plus sign), LOAD DATASET, DATA, DICTIONARY, ANALYSIS (LOAD ANALYSIS, USE TEMPLATE), SAVE, SAVE AS TEMPLATE, MODULE CREATOR.
- Document Area:** Contains a "DATASET Example Dataset" section and a "Table of contents" section with a "show paragraphs" checkbox.
- Bottom Right Panel:** Document recovery (time is GMT+0):
 - 2024-03-24 20:06:37 GMT
 - 2024-03-24 19:48:51 GMT
 - 2024-03-24 18:47:23 GMTComments section.

EDA module example 2



The screenshot shows the FAIR Data Services interface with the NTK Analysis App selected. The top navigation bar includes links for AD Workbench, ntk-datahackathon, FAIR Data Services, AD Connect, a search bar, and a user profile for QANA.FLOREAN. The main workspace is titled "NTK Analysis App" and contains a "Document" tab. The toolbar includes buttons for LOAD DATASET, DATA, DICTIONARY, ANALYSIS, SAVE, and USE TEMPLATE. The right side of the interface features a "MODULE CREATOR" section with checkboxes for "Display module configuration" and "Maximize workspace". A "Table of contents" section is visible on the left, and a "Comments" section is at the bottom right.

Hypothesis testing module



The screenshot shows the NTK Analysis App interface. At the top, there's a navigation bar with links to AD Workbench, ntk-datahackathon, FAIR Data Services, AD Connect, a search bar, and a help icon. Below the navigation is a toolbar with Home, Activity, Files, Database, Tools, Apps, Upload, and Administer buttons. The main area is titled "NTK Analysis App" and contains a "Document" tab with a plus sign icon. A sub-menu bar includes "DATASET Example Dataset", "ANALYSIS", and "MODULE CREATOR" sections with various buttons like LOAD DATASET, DATA, DICTIONARY, LOAD ANALYSIS, USE TEMPLATE, SAVE, and SAVE AS TEMPLATE. There are also settings for "Display module configuration" and "Maximize workspace". On the left, there's a "Table of contents" section with a checkbox for "show paragraphs". The central workspace has a placeholder text "Let's conduct an analysis!". In the bottom right corner of the workspace, there's a "Document recovery" panel showing a log entry: "Recover (time is GMT+0): 2024-03-30 19:13:16 GMT".



Achievements

- The report app with 40 modules for data analysis
- R package for extending Shiny functions
- A hackathon and a scientific publication
- Happy researchers and new collaborations



Interactive tools

- Streamlit- Python; quick dashboard prototype
- Dash- Python; production-ready dashboards
- Jupyter- tech users, education
- Voila- Jupyter notebooks for non-tech users
- Tableau, Power BI- business intelligence



Thank you!



What is the difference between ML and AI?



What is the difference between ML and AI?

ML is written in python/R

AI in powerpoint

