

Práctica: Implementación en Shiny

Introducción

Esta práctica tiene como objetivo practicar todo lo aprendido en Shiny en un proyecto real que además os pueda servir de portfolio de trabajo para vuestra carrera profesional.

Es un esfuerzo importante tanto para vosotros como para la corrección, aprovechad esta oportunidad tan explícita de mejorar vuestro portfolio. En todas las ediciones otros compañeros han usado la práctica en sus entrevistas o charlas. Ha resultado ser muy útil y por ese motivo se mantiene e impulsa este ejercicio.

Método de selección de práctica

Cada grupo de estudiantes (1-5 personas) debe tener un práctica diferente. Puede ser una de las prácticas planteadas aquí en este documento u otra opción que queráis plantear vosotros. Si decidís hacer una práctica vosotros tenéis que **previamente confirmar con el profesor que es válida para hacerla**. También podéis hacer modificaciones a las propuestas que os envío.

Debéis enviar un email a alejandro.vidal@cunef.edu mencionando: los miembros del equipo y las 3 prácticas que queréis realizar en orden de preferencia, sean vuestras ideas o alguna de las recogidas en el documento. Si algún equipo quiere realizar su propia idea puede enviar sólo esa (en el caso de que sea su primera y única opción) y si es adecuada se le asignará directamente.

El 29 de diciembre os contestaré con la práctica asignada si habéis enviado el correo. Si por algún motivo no os es posible decidir práctica antes de esa fecha podría asignarse una práctica posteriormente (pero sin repetir alguna que ya se haya asignado a vuestros compañeros)

A continuación se hace una breve descripción escrita de cada posible práctica y se adjunta unos bosquejos en algunas de ellas de manera orientativa. Si no hay un bosquejo el estudiante puede decidir cómo quiere hacerlo visualmente.

Entrega de la práctica

Se realizará a través del campus mediante la actividad creada para la práctica. La fecha de entrega es hasta el 11 de febrero.

Se deberá entregar una carpeta comprimida con todo el proyecto de RStudio. Recordad que si la práctica requiere recursos externos (datasets, imágenes) estas deben estar contenidas en el entregable.

Evaluación

Hay tres tipos de prácticas: avanzada, media y normal.

- Avanzada: puntúa sobre 11 permitiendo subir la nota final de la asignatura.
- Media: puntúan sobre 10
- Normal: puntúan sobre 8, permiten realizar la práctica más rápido pero limita la nota máxima

Algunas propuestas tienen varios niveles, esto quiere decir que se puede optar a hacer la práctica completa y alcanzar la máxima puntuación o elegir el nivel anterior y optar a una menor nota.

En el caso de sospecha de plagio se podrá solicitar una evaluación posterior con los miembros del equipo.

Seguimiento y dudas

Para la resolución de dudas se deberá reservar tutoría mediante el (siguiente enlace) [<https://calendly.com/alexvidal/tutorias-visualizacion-de-datos-cunef>]:

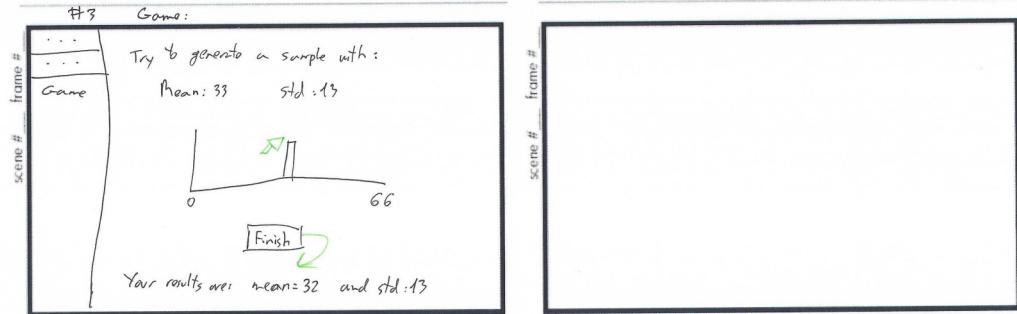
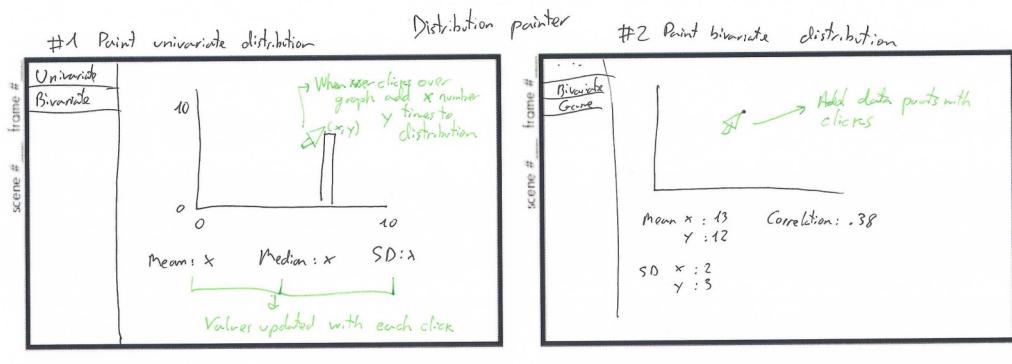
Posibles prácticas

Distribution Painter (Medio y Avanzado)

Se trata de una herramienta de analítica visual y docencia que mediante interacciones de clicks permite al usuario crear sus propias formas de distribuciones y obtener estadísticos básicos.

Existe la versión univariada (histograma) y univariada.

También se añade un juego que consiste en crear una distribución que cumpla unos estadísticos descriptivos proporcionados por la aplicación



Distribution Painter

Probabilistic Excel (Avanzado)

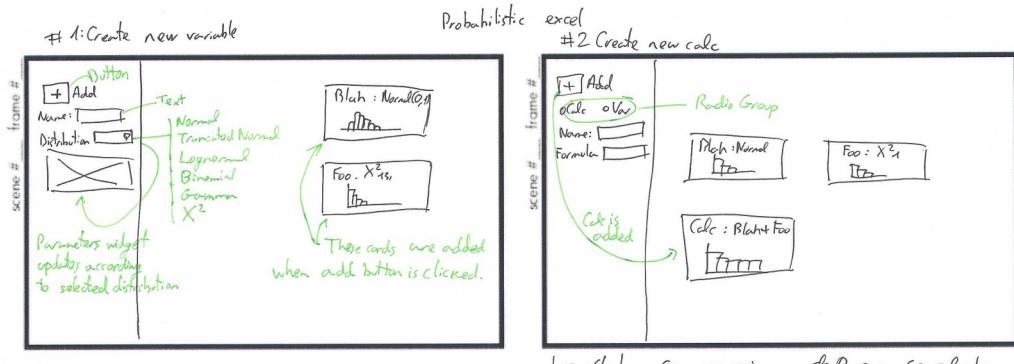
Inspirado en [Caladis](#). Se trata de un “excel” u hoja de cálculos probabilística. De manera que se le permite al usuario hacer cálculos inciertos sobre distribuciones. p.e:

Ventas = N(1000,150) Beneficio por venta = X2(3)

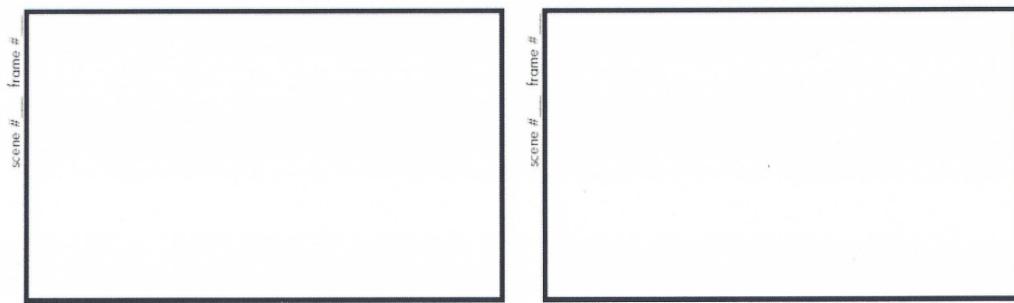
Costes = N(800,13) Beneficios = Ventas * Beneficio por venta - Coste

Se utiliza método monte carlo (simulaciones) para calcular los valores de la distribución.

Cada variable y cálculo se acompaña de una tarjeta con su distribución



Los cálculos son expresiones de R que se calculan con las variables generadas anteriormente.



Excel

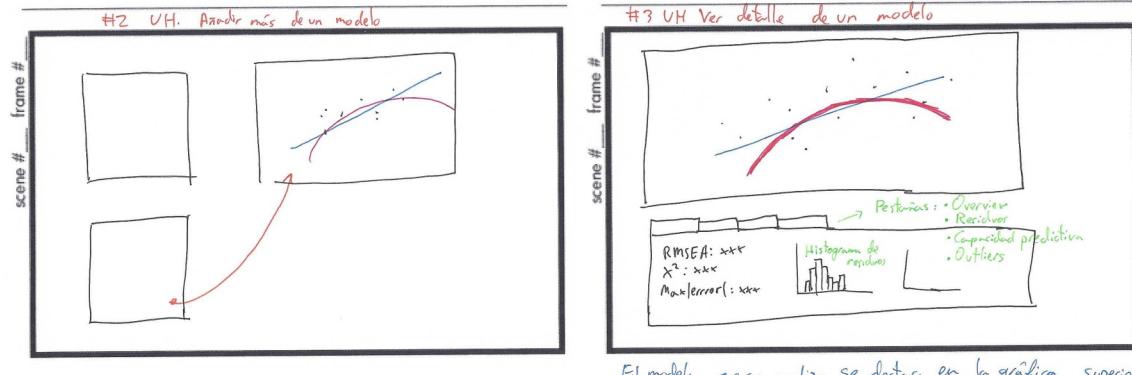
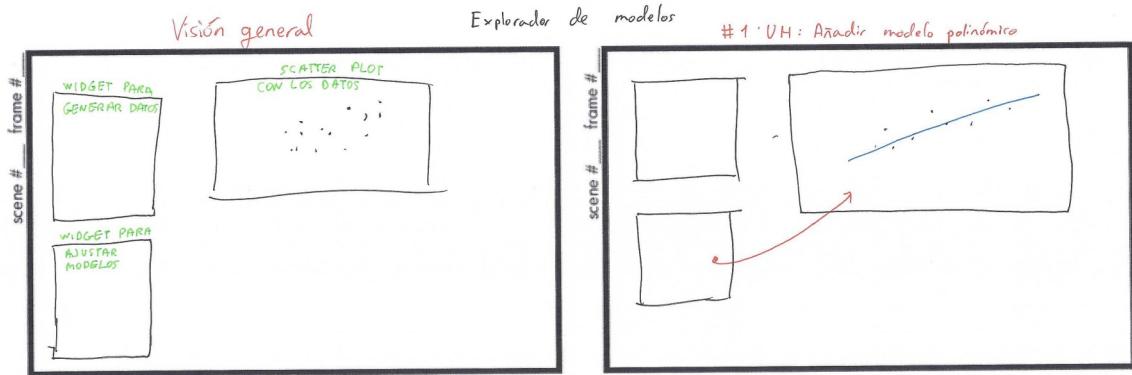
Explorador de modelos (Avanzado)

Herramienta que permite añadir modelos polinómicos a datos arbitrarios

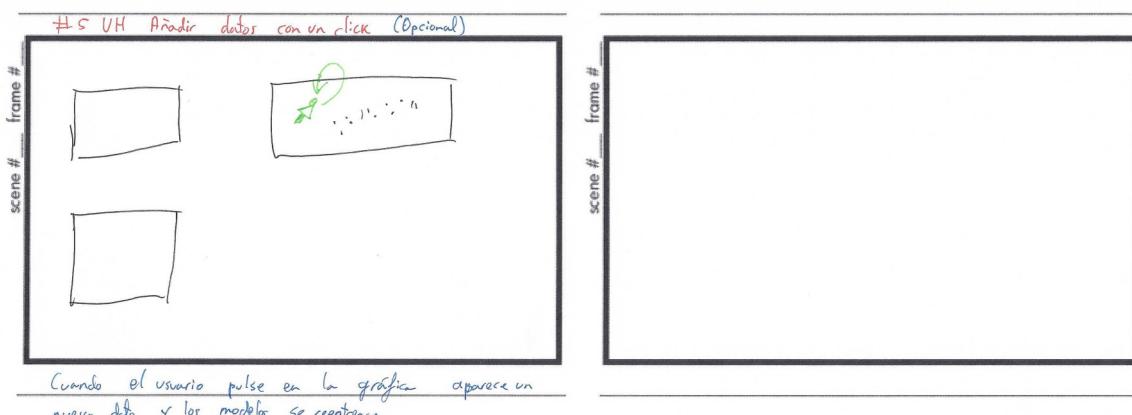
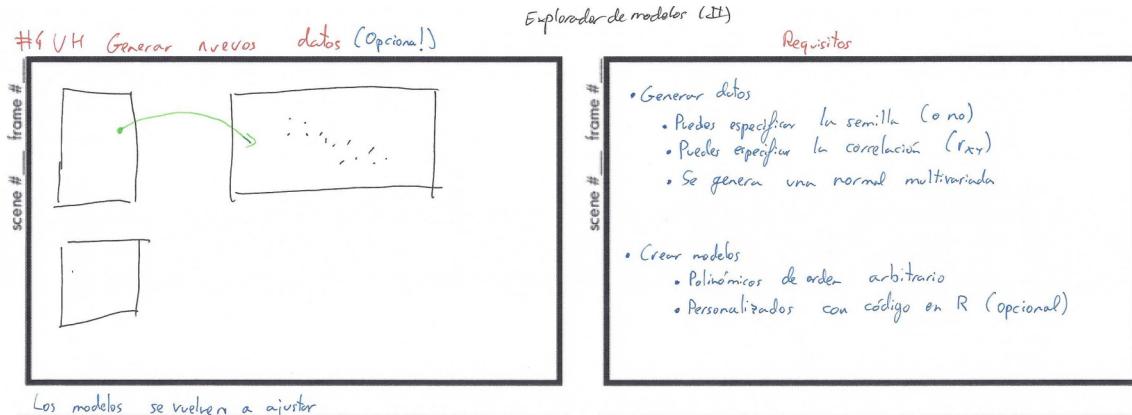
El usuario puede elegir la semilla y la correlación de una normal multivariada. Además pulsando sobre la gráfica puede añadir nuevos puntos de datos.

Todos los modelos se muestran en la misma gráfica junto a los datos y cualquier cambio en los datos o en los modelos hace que se vuelvan a ajustar.

También se añaden unas pestañas con estadísticos de calidad del modelo.



Explorador



Explorador

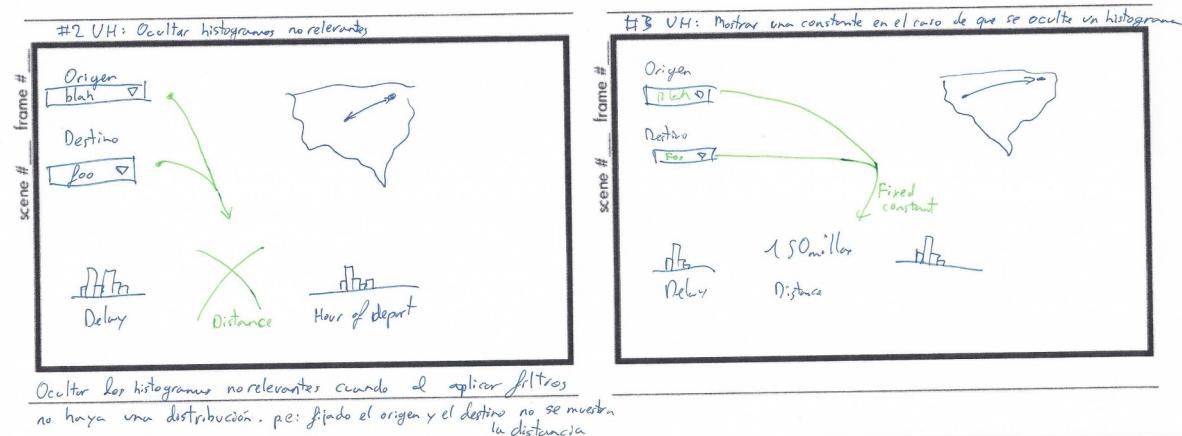
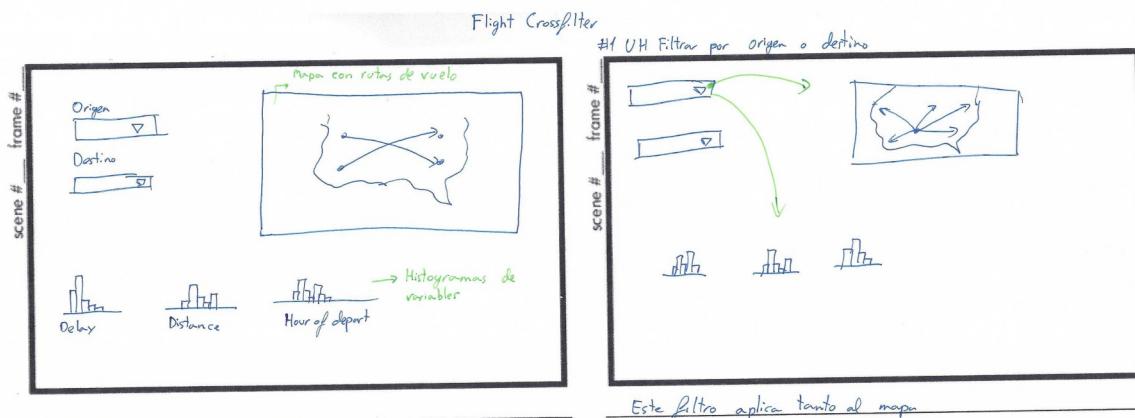
Flight explorer (Medio y Normal)

Herramienta que permite explorar el dataset de delays de vuelos con dos filtros por Origen y destino.

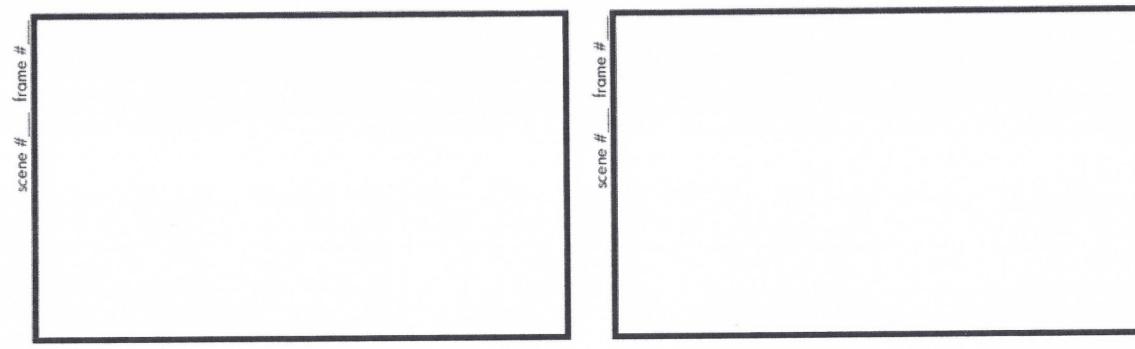
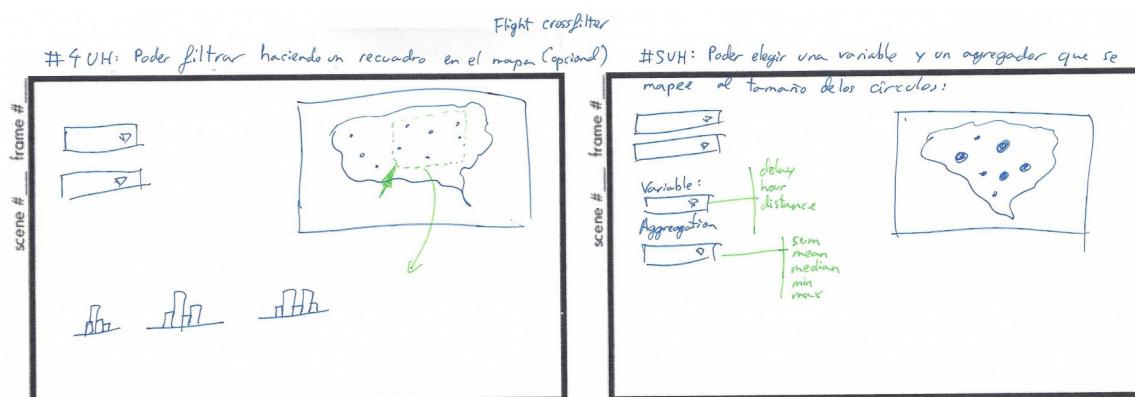
Se muestran histogramas de varias variables y en el caso de que los filtros elijan origen y destino las variables

constantes (p.e: distancia) se muestran como un número constante en vez de una distribución.

También se puede filtrar haciendo un recuadro en el mapa. Además se puede elegir el mapeo de otra variable alternativa y de una función agregadora para que el tamaño de los puntos de los aeropuertos cambie.



Flight



Flight

Function explorer (Avanzado y Medio)

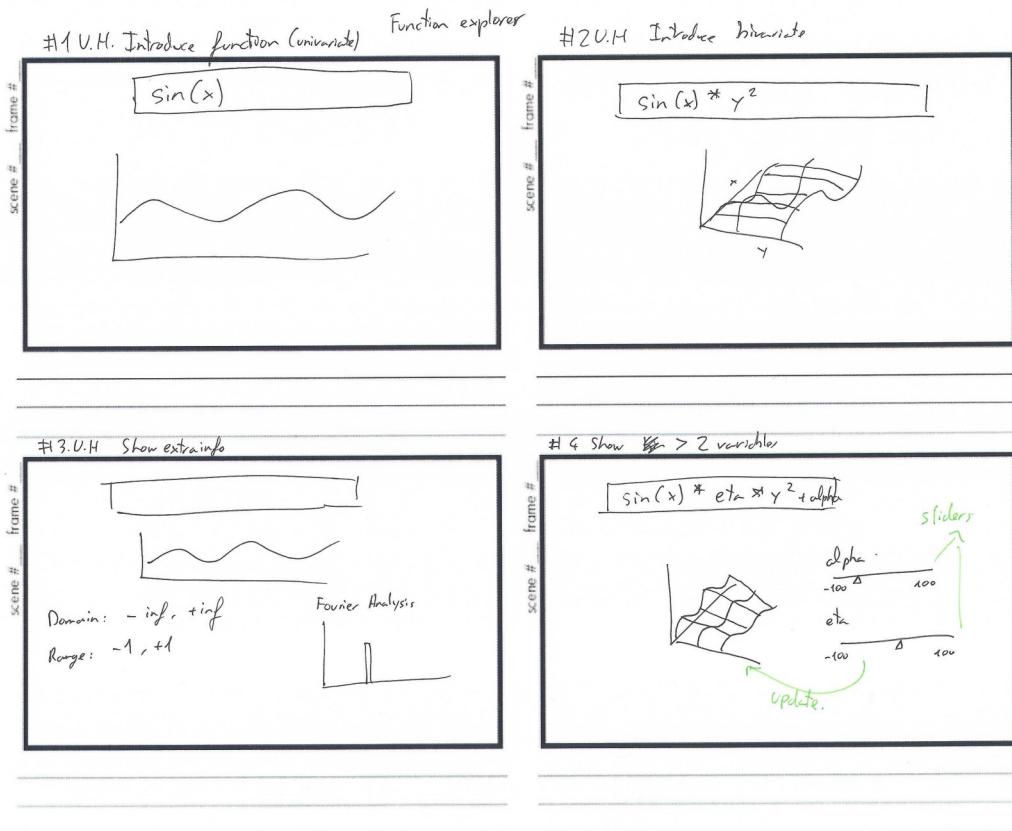
Aplicación interactiva que permite explorar funciones n-dimensionales. Permite cualquier expresión de R arbitraria.

Las variables de la función será cualquier expresión no conocida por R (p.e: x,y, eta o alpha en los ejemplos)

Según sea unidimensional o de dos dimensiones o más se dibujará una gráfica de líneas o una superficie.

Además se añade información sobre la transformada de Fourier de la Función, su dominio y su rango.

En el caso de que haya más de dos variables, las restantes (cogiendo aquellas con longitud mayor) se permiten modificar mediante sliders que se pueden animar. De manera que el usuario pueda “navegar” a través de la animación viendo el efecto de un parámetro

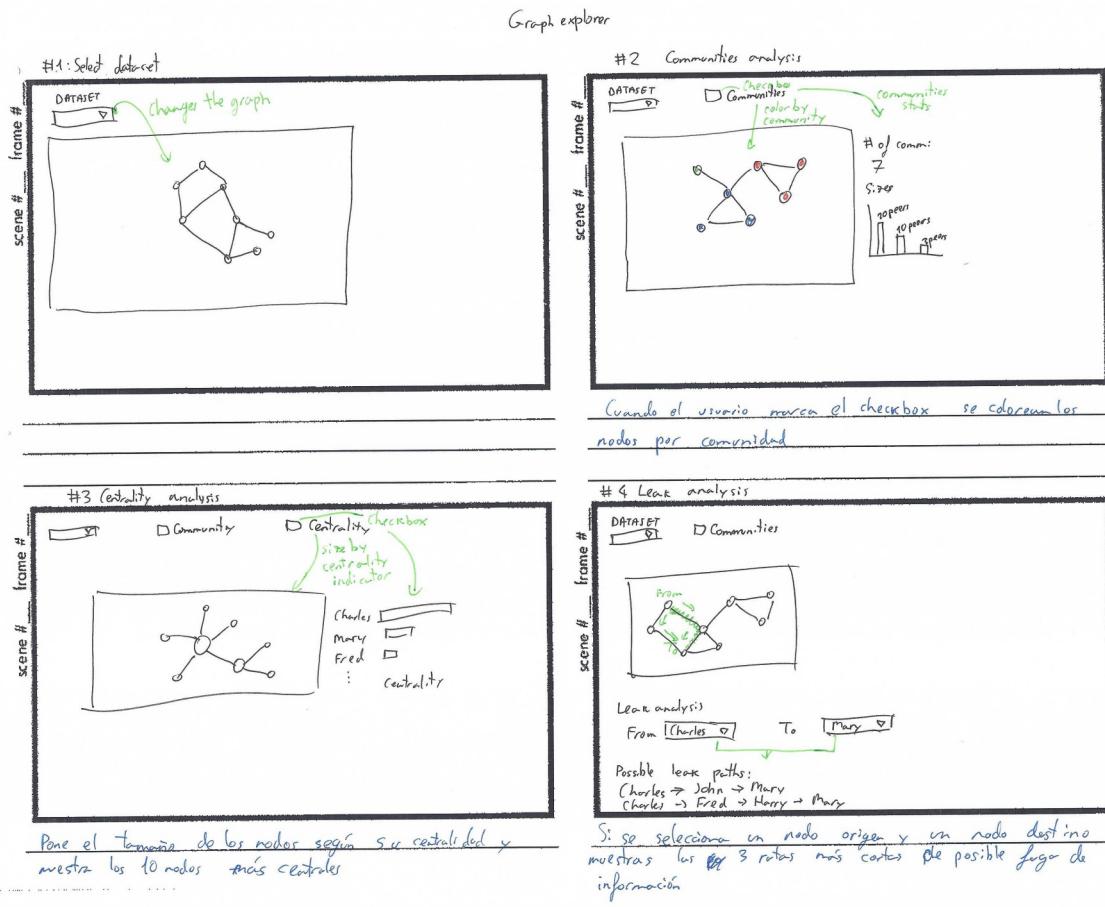


Function Explorer

Graph explorer (Avanzado)

Herramienta de exploración de grafos. Permite elegir distintos datasets y visualizarlos. Para visualizarlos se puede utilizar [networkD3](#) o [visNetwork](#).

El usuario puede realizar un [análisis de comunidades](#), (estudios de centralidad) [<https://en.wikipedia.org/wiki/Centrality>] y posibles rutas de fuga de información entre dos nodos



Graph

Guess the correlation (Avanzado, Medio y Normal)

Inspirado en (Guess the correlation)[<http://guessthecorrelation.com/>]

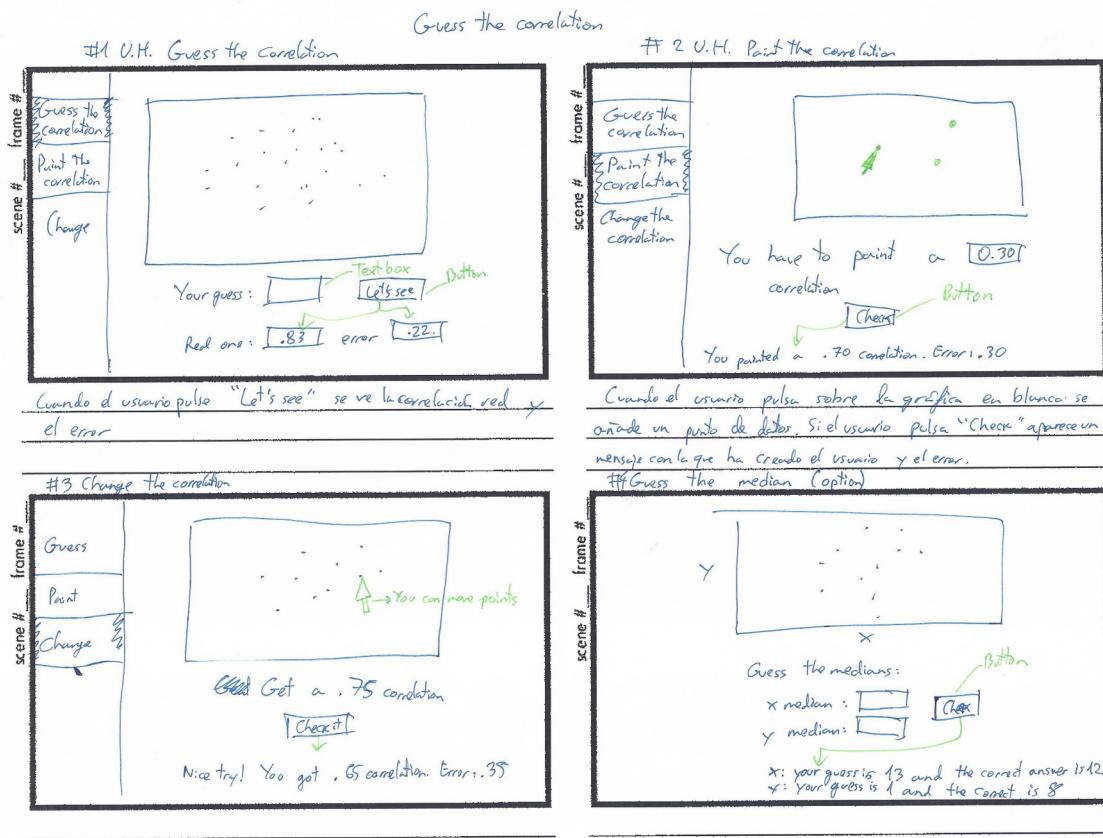
Pequeño conjunto de juegos relacionados con correlaciones.

Primer minijuego de adivinar la correlación de un dataset aleatorio.

Un segundo de pintar un dataset que cumpla la correlación de pearson solicitada.

Otro juego que consiste en modificar un dataset ya existente para cambiar la correlación

Y otro juego que trata de adivinar las medianas de ambas variables



Guess

Simulación de intervalos de confianza (Medio)

Se permite al usuario elegir entre varias distribuciones poblacionales (normal y uniforme) y sus parámetros. Además el usuario elige un número de muestras, el tamaño de la muestra y el nivel de confianza. Se realiza una simulación de un número de muestrear.

Tras pulsar un botón se realiza una simulación de las distribuciones mostrando cuáles de ellas tienen un intervalo de confianza de la media que incluye a la media poblacional.

Se calcula el % de muestras que “aciertan” (el intervalo contiene la media) y que “fallan”. Este porcentaje debería corresponder al intervalo de confianza elegido.

La interfaz debe visualizar los intervalos como se muestra a continuación:

Simulating Confidence Intervals

Describe process

Statistic

Distribution

Method

π

Sample size (n)

Number of intervals

Confidence level

%

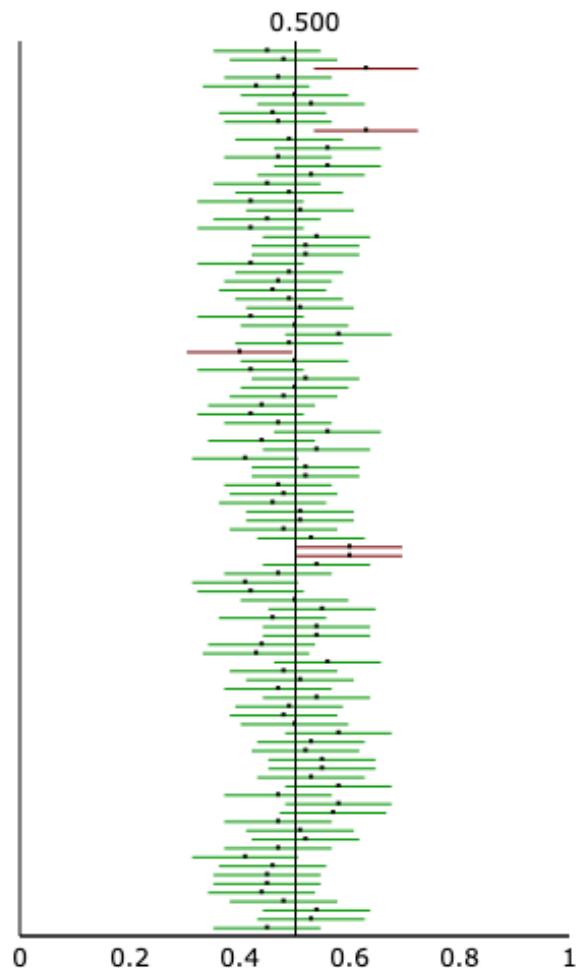
Results

Intervals containing π

$95 / 100 = 95.0\%$

Running total containing π

$107 / 113 = 94.7\%$



Intervalos

[Imagen tomada de este ejemplo](#)

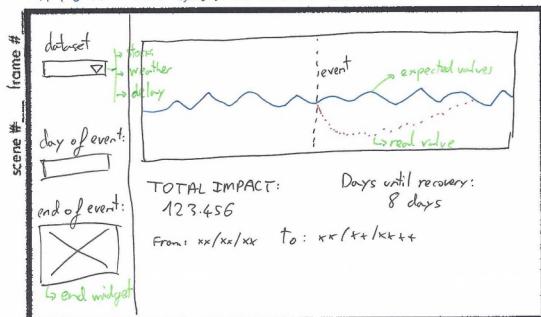
Impact effect model (Medio)

Herramienta visual para utilizar el modelo estadístico (Causal Impact)

[<https://google.github.io/CausalImpact/CausalImpact.html>] con distintos datasets de ejemplo.

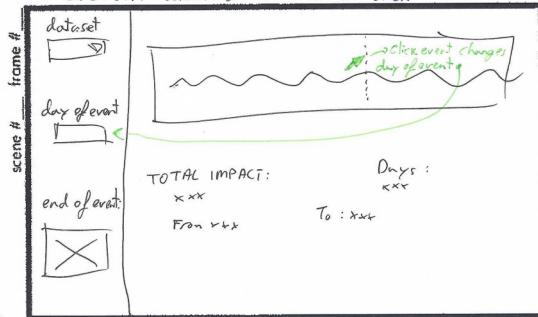
El usuario puede elegir distintos criterios de “parada” del impacto y visualizar distintas métricas (impacto acumulado, puntual y relativo)

#1 U.H: Mostrar estadísticas básicas del evento

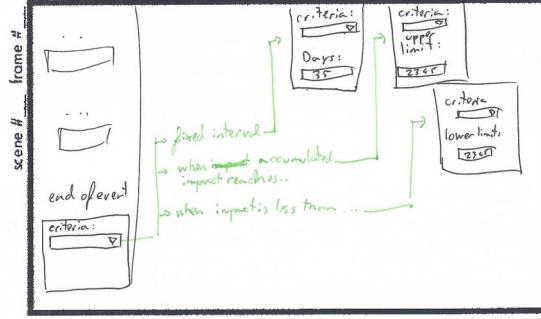


Impact effect model

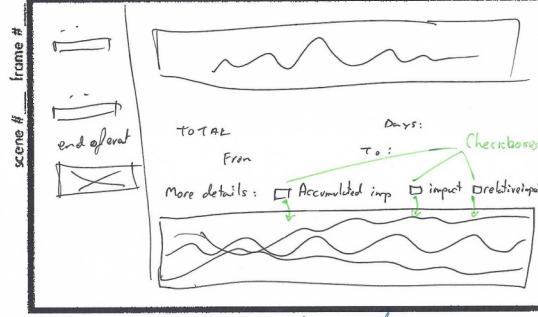
#2.U.H Señalar el evento con un click



#3.U.H: Otra Cambiar el criterio de fin del evento



#4.U.H Ver otros gráficos



El usuario puede marcar que otros gráficas quiere ver pulsando marcar varios simultáneamente.

Impact

Influence Analysis (Avanzado y Medio)

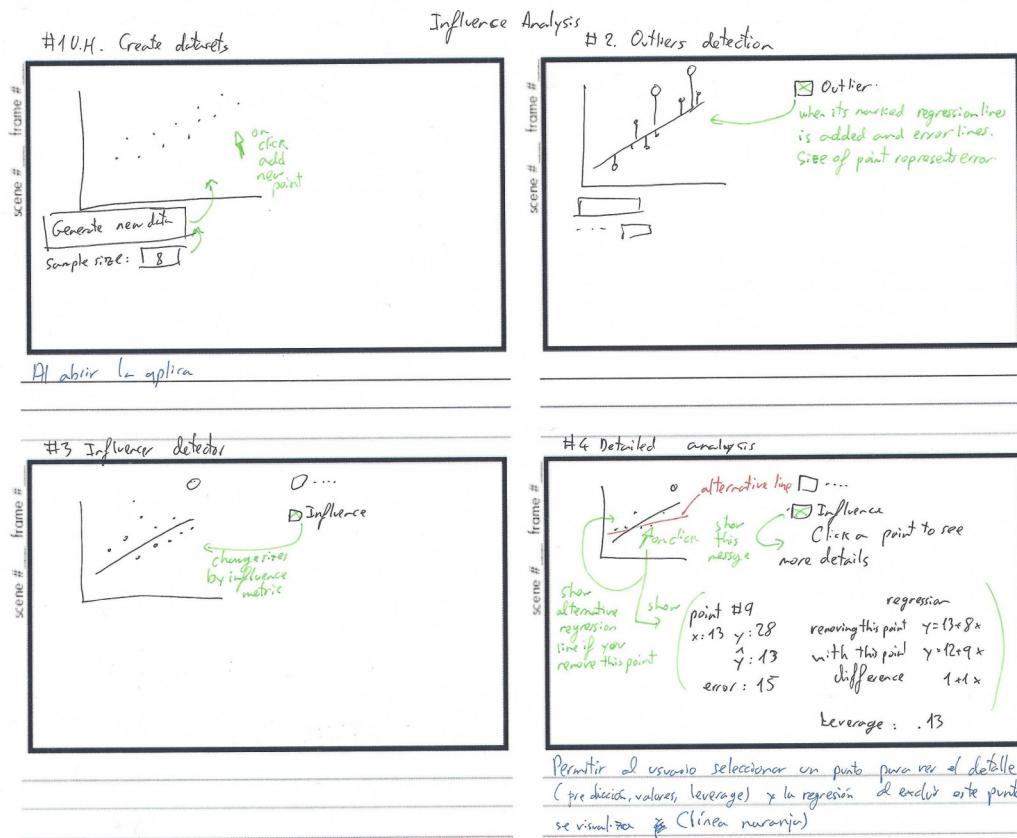
Herramienta de Visual Analytics para diferenciar de una muestra bivariada aquellos datos que son outliers (con alto error) de aquellos que tienen mucha influencia en la regresión.

El usuario puede generar una muestra al azar del tamaño seleccionado.

Hay una visión de outliers en la que se muestra la línea de regresión y de errores fijos. También se codifica en tamaño el error de estimación de cada punto.

Hay una visión de influencia en la cual el tamaño del círculo se mapea al indicador de influencia y se calculan las regresiones de toda la muestra y la muestra excluyendo este punto.

Ambas regresiones se superponen a los datos y se comparan.



Influence

Lotto bias (Avanzado y Medio)

Herramienta para estudiar los sesgos perceptivos de azar mediante un juego de loterías. Para ello se utilizan los datos reales de loterías públicas. Se compone de 3 minijuegos:

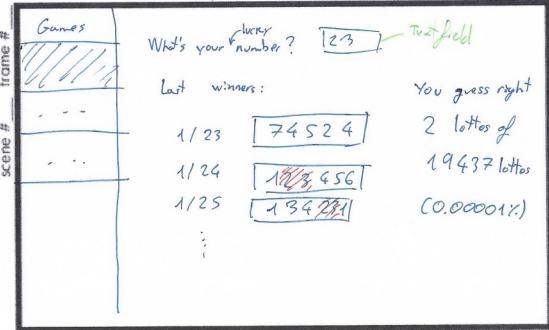
1. What's your favourite number?: El usuario selecciona un número y se revisa en el histórico real de loterías cuántas veces ha aparecido ese número mostrando la estadística al usuario
2. Guess the good one: Se muestra un número del dataset real de loterías y otro generado al azar por R. Existe una creencia a que los números que tocan en la lotería cumplen ciertos patrones (repetitividad, recencia negativa, no empiezan por cero, etc.). Como son totalmente aleatorios el sujeto no debería poder distinguirlos entre ambos. Las posiciones del número "real" y el aleatorio se sortean cada vez que el usuario responde.
3. De un dataset "human-trap numbers" (números "bonitos" o aparentemente más probables) se selecciona dos números y se añaden una opción más realmente aleatorias. El usuario debe seleccionar aquella que está diseñada para atraer la atención ("human-trap")

Lotto bias.

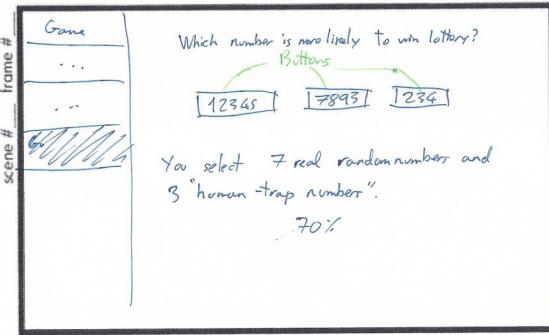
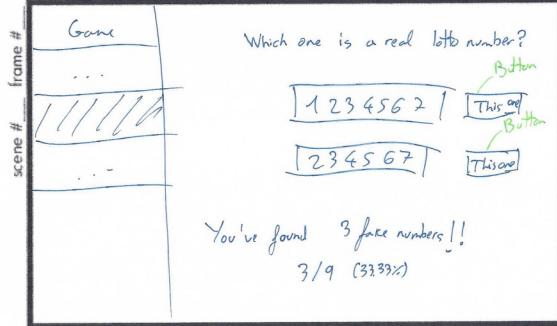


El usuario tiene un menú para seleccionar a que juego quiere jugar.

#1 VH: What's your favourite number



#2 VH: Guess the good one



Influence

KMeans analytics (Medio y normal)

Permite ejecutar KMeans con un dataset a elegir por el usuario entre varias opciones y elegir las variables que se utilizan.

Visualiza en 2d y 3d (según corresponda) el resultado del KMeans

Opcionalmente muestra el solapamiento entre cluesters de distintas ejecuciones del Kmeans. Es decir que % coincide entre clusters de un KMenas de 3 clusters y de 4 clusters.

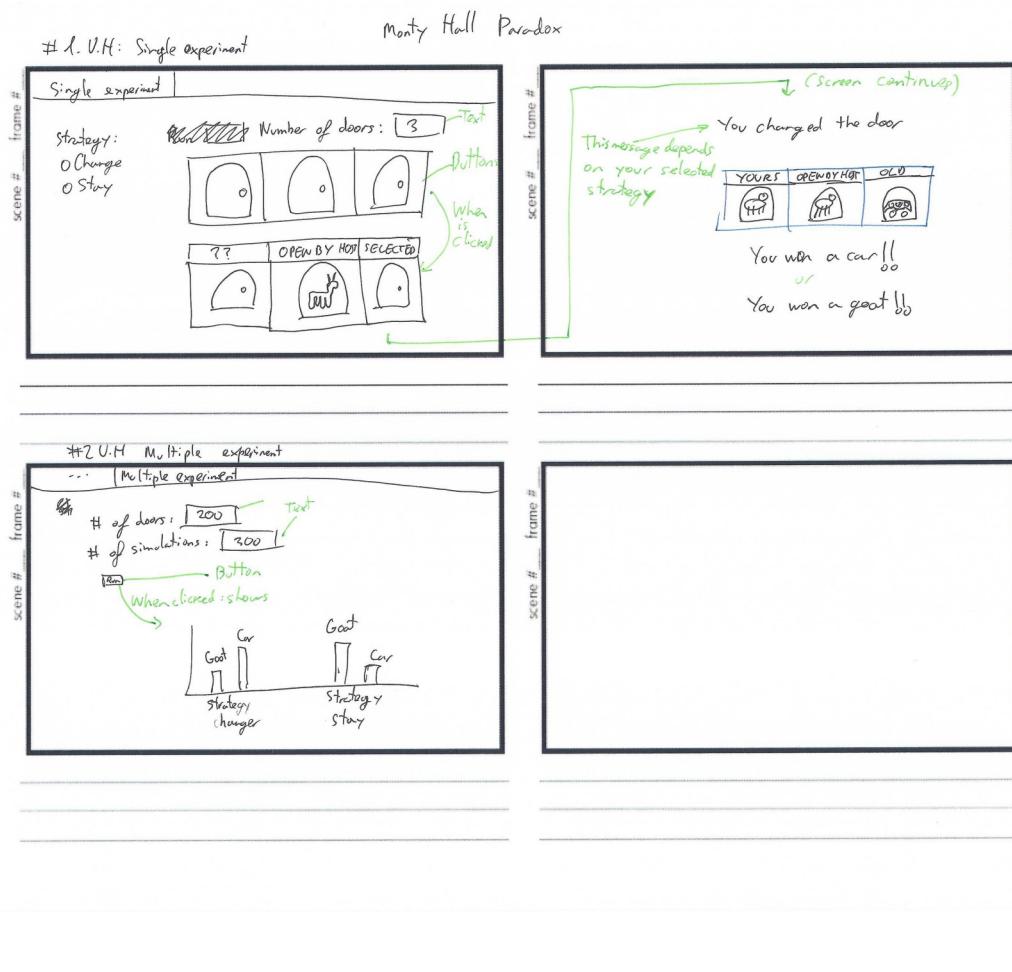
Monty hall paradox (Medio y Avanzado)

La paradoja de Monty Hall es una contraintuición muy popular. Ver funcionamiento del juego en [Problema de Monty Hall \(Wikipedia\)](#).

Se trata de reproducir cómo es el juego bajo dos estrategias de juego distintas: mantener la puerta y cambiar la puerta (leer el artículo de la Wikipedia).

El programa permite hacer simulaciones simples para enseñar el funcionamiento al usuario y cambiar el número de puertas (esta es una manera frecuente de enseñar esta paradoja).

Otro apartado de la aplicación permite ejecutar múltiples simulaciones para comprobar que la estrategia óptima es cambiar de puerta.



Influence

Simple calculator (Normal)

Permite hacer operaciones aritméticas sencillas con dos operando (dos números) mediante botones. La calculadora puede realizar: sumas, restas, divisiones y multiplicaciones.

Buscador de libros con velocidad de lectura (Normal y Medio)

Interfaz simple que permite buscar ISBN para encontrar el título y otras propiedades del libro. Adicionalmente permite calcular la velocidad de lectura del usuario y estimar cuento tiempo de lectura para ese libro sería necesario.

Interfaz para la evaluación de alumnos (Normal)

Una interfaz que carga un fichero CSV con las notas de varios alumnos. La interfaz permite modificar el resultado de las actividades y detectar alumnos que destaque (outliers) por alto o bajo rendimiento.

Perceptual experiment (Avanzado)

Utilizando las imágenes de MNIST (-ver Wikipedia-) [https://en.wikipedia.org/wiki/MNIST_database] esta aplicación permite diseñar y ejecutar un diseño experimental factorial.

La tarea a resolver es el reconocimiento del dígito mostrado la imagen original con un poco de ruido. El experimento trata de analizar el efecto del ruido en la percepción de los dígitos.

Para eso hay varios tipos de ruido: - Ruido blanco: en puntos aleatorios de la imagen se coge ruido de una distribución aleatoria - Zeroing: En puntos aleatorios se pone a 0 la casilla (se pone a negro) - Vertical lines: se añaden líneas verticales negras a la imagen al azar - Small squares: en regiones cuadradas al azar de un tamaño determinado se tapa la imagen

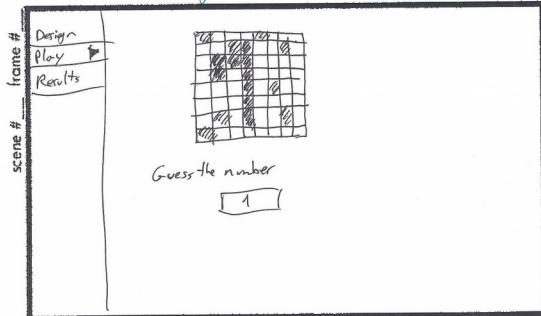
La aplicación muestra todas las combinaciones según se hayan activado tipos de ruido. (Es un diseño factorial, por tanto será multiplicar el número de alternativas. p.e: tipos de ruido* porcentajes de ruido*...)

En la sección play el usuario puede realizar el experimento. Se debe permitir que más de un usuario haga el experimento después.

En la sección results se pueden ver los resultados (en forma alargada).

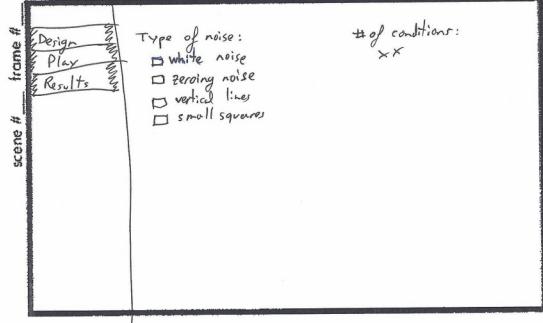
Perceptual experiment

#1 U.H. Play the game

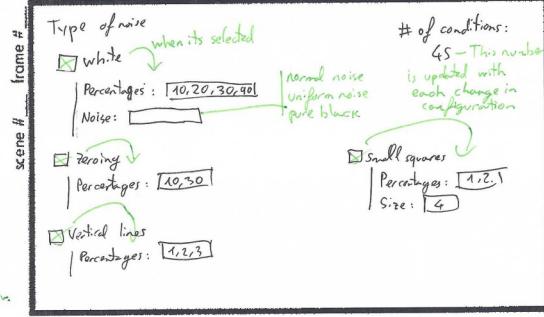


En la pantalla "Play" el usuario tiene una gráfica con un ruido de punter.

#2 U.H. Design the experiment: select noises



#3 U.H: Select levels of noise



#3 U.H See results

Results	Type of noise	Percentage	Noise	Subtype	Red number	Position	Position inside experiment
TYPE
...
...

perceptual

P hacking (Medio y Avanzado)

El p-hacking es la práctica ilegítima de modificar el diseño de un análisis o los datos con el fin de obtener un resultado estadísticamente significativo.

Muchos expertos apuntan a que estas prácticas se ven fomentadas por el sesgo de publicación de las revistas científicas que sólo suelen publicar resultados significativos aunque eso no es siempre relevante.

Esta aplicación permite mostrar distintas estrategias para manipular los valores p de una correlación:

- Subsample: coger una muestra más pequeña hasta que obtengamos un resultado significativo
- Weighting: ponderar los casos por una tercera variable
- Change variables: elegir otras variables del dataset que correlacionen más

P hacking

#1 : Subsample trick

Select your trick:
Subsample
Weighting
Change variables

Detailed menu

Scatterplot
Correlation $r = 0.73$ p-value 0.03
Linear regression line

#1 : Subsample trick

Select your trick:
Subsample
Weighting
Change variables

Subsample size: 13 - text field

Random - Button

Original scatter
Subsample Correlation $r = 0.92$ p-value 0.13

En esta aplicación se dibuja una población bivariada en un scatterplot, la correlación de pearson y el p-value asociado

Cuando seleccionas una opción en el menú la parte de detailed menu (cuadro vacío) aparecen varias opciones. Cuando se pulse el botón se elige una muestra al azar y se pinta en la segunda gráfica.

#1/H3: Weighting

Select your trick:
Weighting

weight variable: others variables in dataset

Original
Weighted Correlation $r = .12$ p-value 0.03

#1/H3: Weighting

Select your trick:
...
Change variables

Select X
Select Y

Original
Never Correlation $r = .78$ p-value 0.05

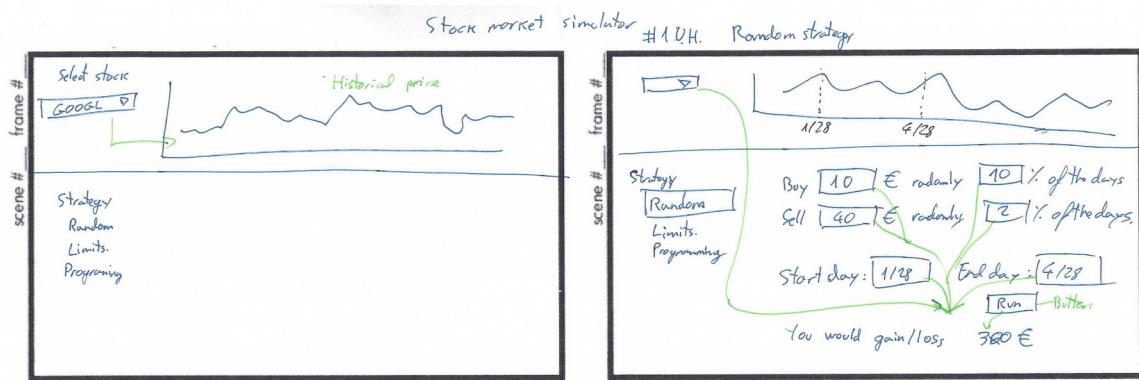
Cuando el usuario elige una variable para pesar los datos se muestran un scatterplot en el que el tamaño de los círculos y la regresión se actualizan con esos pesos

phacking

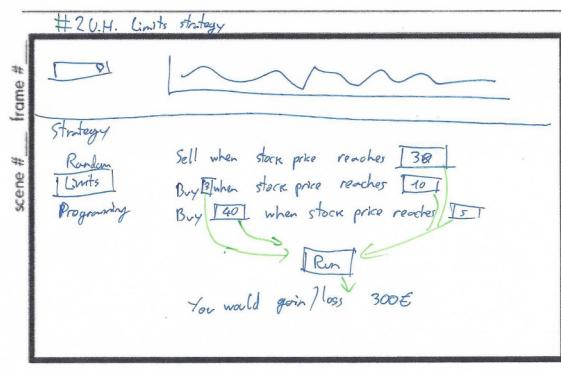
Stock (Avanzado, Medio y Normal)

Aplicación para estudiar el rendimiento contra histórico de valores de stock.

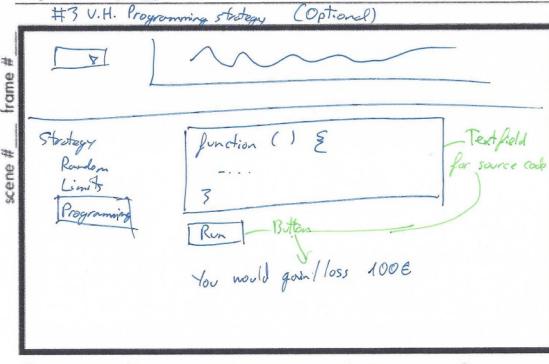
Hay tres estrategias: - Random: compra y venta al azar de una cantidad fija. - Limits: comprar y vender al alcanzar un límite superior e inferior - Programmatic: puedes programar tu propio algoritmo de trading en R y evaluarlo



Cuando el usuario cambia en el selector el stock la grafica del histórico se actualiza

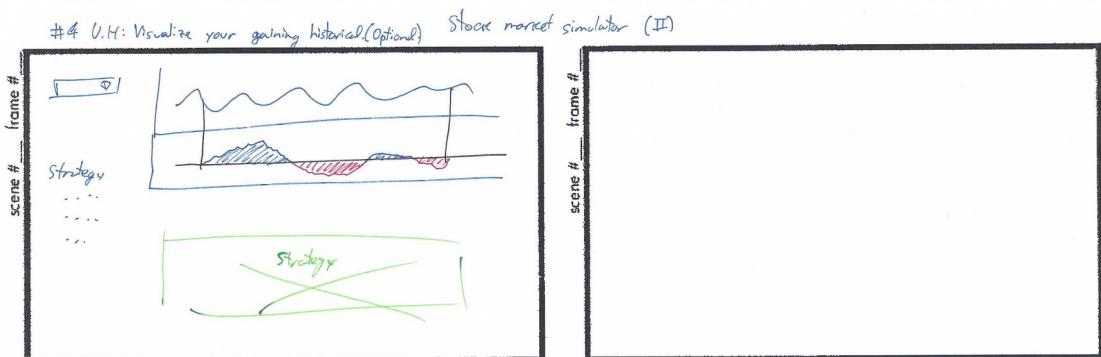


Cuando el usuario pulse en "Run" ejecutaremos una simulación de cuanto hubieras ganado con esa estrategia. También cuando el usuario cambia de stock



El usuario puede programar su propia estrategia en R

stock



stock2

Entrenamiento de mecanografía (Medio y Normal)

Herramienta para entrenar la calidad de la mecanografía del usuario. Muestra una frase y permite al usuario escribirla. A medida que escribe aparece nuevas frases hasta llegar a un límite.

Informa de estadísticas básicas de errores tipográficos y de velocidad de escritura además de la evolución

Birthday paradox (Avanzado)

La paradoja del cumpleaños no es realmente una paradoja sino una conclusión matemática contraintuitiva.

La paradoja habla sobre la probabilidad de que dos personas cumplan el mismo día del año en un mismo grupo de personas (normalmente una clase de estadística).

Con sólo 23 personas la probabilidad ya es superior al 50% y con 47 es superior al 90%.

Esta aplicación permite visualizar esta paradoja con una visión de calendario y ejecutando una simulación en la que se van añadiendo sujetos al grupo.

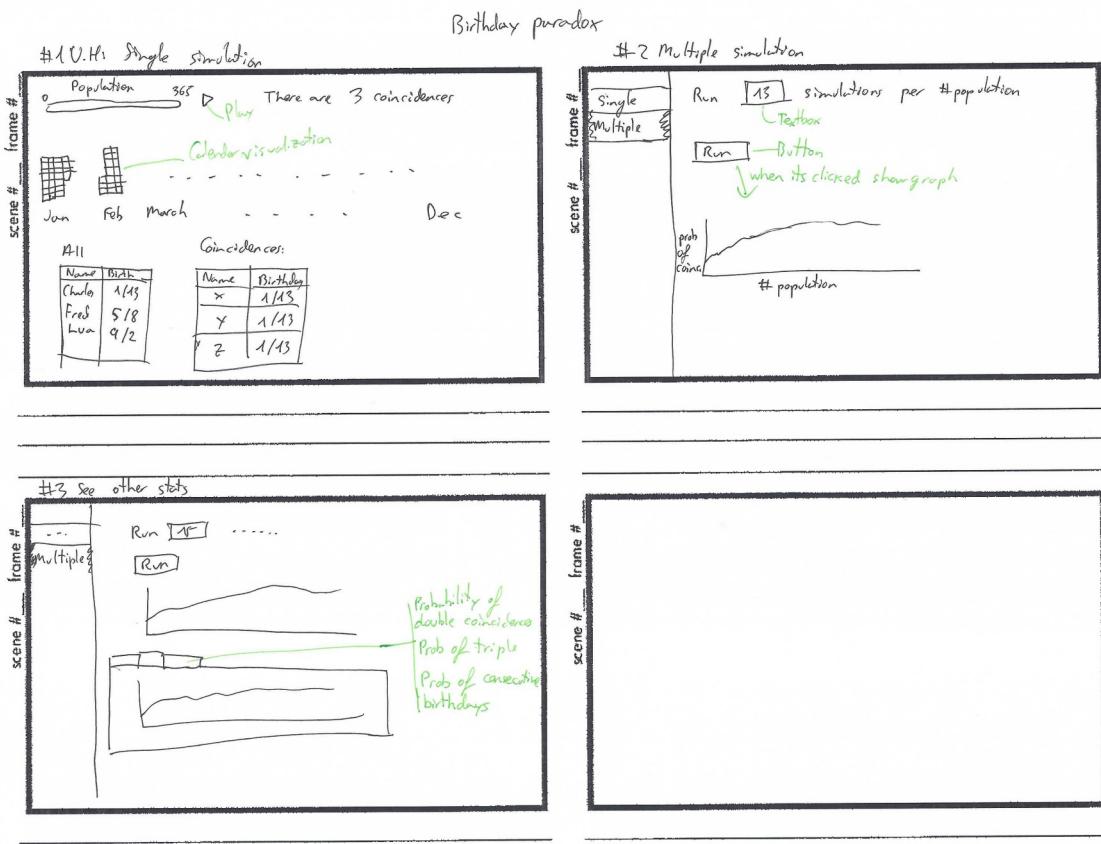
El usuario puede controlar el tamaño de la clase.

Además se puede ejecutar una simulación masiva para calcular mediante una aproximación monte carlo la probabilidad real de coincidencia.

Por último se analiza la probabilidad de otros fenómenos “llamativos” como la probabilidad de doble coincidencia (tres personas cumplen el mismo día), triple (cuatro personas) y cumplir años en días adyacentes.

Para la visualización del calendario se puede usar [rChartsCalMap](#), [calheatmapR](#) o [googleVis](#).

Más información: [Wikipedia](#)



Birthday

MNIST Machine Learning tester (avanzado)

Interfaz que te permite comparar el rendimiento de distintos modelos (regresión lineal, SVM, ...) en el dataset de MNIST.

Adicionalmente permite obtener el resultado en una imagen en concreto y la modificación de esta imagen por el usuario pintando sobre ella para ver como afecta sobre el resultado para cada modelo

Visualizador de mapas de ciudades (Normal)

Una aplicación de shiny que te muestra el mapa de varias ciudades además de algunas propiedades extraídas de un dataset demográfico oficial.

Hay que usar una librería para la visualización del mapa y geolocalizar correctamente los mapas

Selector de polinomio óptimo para un modelo (Medio y Normal)

Herramienta para elegir la complejidad de un modelo lineal necesaria mediante el estudio del hiperparámetro orden: lineal, cuadrático, cúbico, ...

Para ello se permite al usuario navegar entre las distintas opciones dado un dataset cargado con CSV o un dataset aleatorio.

Además se le muestra una gráfica con la relación entre % de varianza explicada (o error) y grado. De manera que se pueda analizar hasta qué punto merece la pena hacer más complejo el modelo.

Opcionalmente se puede hacer esto para training y test.