rmed\_googlesheet\_abstracts

Peter Higgins

7/13/2021

## Read in the abstracts

This assumes that you have tidyverse, glue, and googlesheets4 installed. You will have to authorize access to your google drive account for the first use.

abstracts <- read\_sheet("https://docs.google.com/spreadsheets/d/1CJcH7vCw2nUcrQ\_79LNfTPsE0rN-p-\_afyO9hMOTutE/edit") %>%   
 slice(1:64)

## → Using an auto-discovered, cached token

## To suppress this message, modify your code or options to  
## clearly consent to the use of a cached token

## See gargle's "Non-interactive auth" vignette for more  
## details:

## <https://gargle.r-lib.org/articles/non-interactive-auth.html>

## → The googlesheets4 package is using a cached token for  
## 'pdr.higgins@gmail.com'

## Reading from "abstract-7-13-2021"

## Range "abstract-7-13-2021"

## Printing info

We will use glue\_data from the glue package

## name: Mihir N. Patel   
## org: Duke University School of Medicine  
## title: Medical Student (MS1)   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I am a first-year medical student at the Duke University School of Medicine and completed my undergraduate education at Duke University in Durham, NC. I have worked with Dr. Arif Kamal and his team at the Duke Cancer Institute over the previous two years. Through our research, I have developed a passion for developing mobile applications to address quality and patient education gaps in hospice and palliative care delivery and leveraging big data (such as from search engines and social media platforms) to understand trends and barriers in hospice and palliative care utilization.  
## gender: male  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: Using Google Trends to understand search interest patterns for hospice and palliative care  
## abstract: Nearly half of all Americans who die never receive hospice care. Misunderstandings and unfamiliarity with hospice among patients impede use. Knowledge regarding information-seeking patterns of patients and caregivers may inform interventions to increase uptake. Internet search activity is a reliable indicator of public interest. Google Trends provides spatiotemporal Google search activity for any search term and is accessible via the gtrendsR package. We modified an existing data collection protocol to obtain relative search volumes (RSVs) for hospice and palliative care search terms across all U.S. states and D.C. from 2015 to 2018. We found that peaks in general public search interest corresponded to media coverage of hospice use by notable celebrities. Using choropleth maps, we identified states where inaccessibility of hospice and palliative care programs corresponded to elevated searches for hospice and palliative care. After pooling state-level hospice use and demographic data from multiple sources, we used a generalized estimation equation model to show that searching for “hospice” is associated with earlier hospice use. Our work underscores the value of search activity data in elucidating hospice interest and use patterns.  
##   
##   
## Your Score (1-9):   
##   
## name: Nathaniel Fernandes   
## org: Texas A&M University  
## title: Undergraduate Researcher   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Nathaniel Fernandes is a Brockman Scholar at Texas A&M University, pursuing a computer engineering degree. Incredibly curious, he loves to learn and enjoys working with others. Nathaniel plans to apply artificial intelligence to healthcare to guide physicians' decisions and enhance patient-care worldwide. His hobbies include playing chess, programming, and web development (he’s the proud webmaster of chessintellect.com). In addition, he is an Eagle Scout, National Merit Finalist, and graduated salutatorian of his high school class. His favorite saying is, “The more you sweat in practice, the less you bleed in battle.”  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Open-source Algorithm to Calculate Mean Amplitude of Glycemic Excursions Using Short and Long Moving Averages  
## abstract: Background: Glucose variability is a risk factor for diabetes complications. The mean amplitude of glycemic excursions (MAGE) is a classic measure of glucose variability, but it can require manual calculation. Existing programs for automatic MAGE calculation show varying agreement and lack rigorous validation. This study aims to establish a publicly available gold standard which any MAGE algorithm can be compared against, and to develop an automated MAGE algorithm which shows high accuracy against the gold standard. Methods: Manual MAGE calculations were performed on 45 CGM traces from a diverse set of patients (Type 1 diabetes, Type 2, non-diabetic). An algorithm for MAGE calculation has been developed. We estimated the accuracy of the proposed algorithm with a five-fold cross-validation on the samples and compared it to four publicly available software programs for MAGE calculation. Results: The proposed algorithm has the highest accuracy in calculating MAGE with a cross-validation error of 9.0% relative to manual calculations. Conclusions: The proposed algorithm eliminates the need for tedious manual MAGE calculations and is validated to be more accurate than existing algorithms when compared to the manual gold standard.  
##   
##   
## Your Score (1-9):   
##   
## name: Julia Romanowska   
## org: University of Bergen  
## title: senior engineer   
## time zone: (GMT+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna  
## bio: I am a highly creative programmer and a mother of two. I studied physics and informatics at University of Warsaw in Poland, and went on to complete a PhD in computational biophysics. After that, I worked as a post-doctoral research at Heidelberg Institute for Theoretical Studies and at the University of Bergen, bridging bioinformatics and genetic epidemiology. Currently, I am a full-time data scientist at the Core Center for Biostatistics and Data Analysis at University of Bergen.  
##   
## During my career, I have learned how to code in different programming languages, alone and in a team, I have explored various datasets, and created different visualizations; I have been presenting my results to various audience. Moreover, I have teaching experience, successfuly applied for funding for my own research, and thereafter, managed my research projects. My results were highlighted and awarded several times. I am also a co-founder of R-Ladies Bergen. Privately, I love to learn new things and create - I am a knitter and I’ve recently learned to sew, I love nature, yoga, and my small family.  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Diving into registry data: using R for large-scale data analyses on Norwegian health registries  
## abstract: Traditional process of designing new drug is timely, costly (https://en.wikipedia.org/wiki/Cost\_of\_drug\_development), and often fails (https://en.wikipedia.org/wiki/Drug\_development#Clinical\_phase). However, by analyzing the usage patterns of the drugs already on the market, alongside the history of illnesses of each person, new drug application areas might be discovered - so-called drug repurposing.  
## In Norway, national registries have been collecting various types of data on each Norwegian citizen for many decades. Within our project, (Drug Repurposing for NEurological diseases, http://link.uib.no/drone), we have access to the entire Prescription Registry, along with selected diagnoses from Patient Registry, and demographic data from Statistics Norway. This huge amount of information requires appropriate programming approaches so that the full potential of the datasets is harnessed. We have chosen R, with its vast number of packages and ease of coding, to implement advanced methods of epidemiology, bioinformatics, and machine learning to find new drugs for neurological diseases. I will briefly present our research method and point to several R packages and techniques we are using to make our work reproducible and effective.  
##   
##   
## Your Score (1-9):   
##   
## name: Raha Maryam Dastgheyb   
## org: Johns Hopkins Hospital  
## title: Instructor   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I am Junior Faculty at Johns Hopkins Neurology with a background in Biomedical Engineering and Computer Science. I spent most of my career doing wet lab experiments and writing software to analyze them, but in the past few years I have transitioned to collaborating with clinicians to analyze cohort data and search for potential biomarkers of disease.  
## gender: female  
## poc: yes  
## submission\_type: regular  
## abstract\_title: Creating R Pipelines to Uncover, Visualize, and Interpret Latent Clinical Phenotypes  
## abstract: Understanding the mechanisms of disease pathology is the key to being able to treat it. It is important to take a step back and see what the data shows us to make sure we are properly defining the disorder. In this example, we will show data regarding cognitive impairment in HIV, which remains frequent even in the modern era of highly effective antiretrovirals. Traditional definitions of cognitive impairment lump all domains into one “global” score, which we believe washes out granular differences in patterns and lacks sensitivity to uncover potential mechanisms. We will demonstrate implementing a pipeline consisting of dimension reduction and clustering to uncover latent phenotypes of cognitive profiles based on clinical neurocognitive testing data. We will walk through exploring the data, optimizing, and visualizing the clustering. Finally, we will use machine learning models to identify predictors of cluster membership and visualize variable importance using lollipop plots. We treat this as a “hypothesis-generating” model to see what could be contributing to different neurodegenerative mechanisms.  
##   
##   
## Your Score (1-9):   
##   
## name: John Schwenck   
## org: Texas A&M University  
## title: Graduate Researcher   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Having completed his undergraduate education at the Pennsylvania State University in Economics and Supply Chain Information Systems and his graduate education at Texas A&M University in Statistics, John Schwenck has parlayed his statistical modeling and data analytics skills into a wide variety of industries including the energy, sports, finance, and biostatistics sectors. With a passion for Hypertension research, John strives to develop user-friendly software for clinicians and researchers alike with the hope of improving and advancing the medical community.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: bp: Blood Pressure Analysis in R  
## abstract: Despite the world-wide prevalence of Hypertension, there is a lack of open-source software dedicated to analyzing blood pressure. The R package – bp – fills this void by providing functionality for data processing, visualization, and statistical analysis of blood pressure data. The package implements statistical metrics and visuals from the medical literature on Hypertension and is equipped with six open-source sample datasets covering continuous arterial pressure data, ambulatory blood pressure monitoring (ABPM) data, and home blood pressure monitoring (HBPM) data to help researchers get started. This talk will highlight the main bp package functionality and will illustrate a typical analysis workflow that can be used to generate blood pressure reports.  
##   
## The bp package can be accessed via CRAN or the development version on GitHub at https://github.com/johnschwenck/bp  
##   
##   
## Your Score (1-9):   
##   
## name: Ijeamaka Anyene   
## org: Kaiser Permanente Northern California, Division of Research  
## title: Senior Consulting Data Analyst   
## time zone: (GMT-08:00) Pacific Time (US and Canada); Tijuana  
## bio: Ijeamaka Anyene is a Senior Consulting Data Analyst at Kaiser Permanente Northern California, Division of Research where she works on research investigating the links between body composition and cancer survivorship. She has previously received her MPH in Epidemiology and Biostatistics from U.C. Berkeley School of Public Health.  
## gender: female  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: Tracking Dashboards and Specimen Workflow Management using R and {shinydashboard}  
## abstract: In the process of creating cohorts for healthcare research, it is often required to consolidate data from various sources. These sources include medical records, lab and pathology results, and data collection software such as REDCap. However, having disparate data sources requires individuals to be skilled in using multiple software systems to 1) have visibility into the data collected and 2) manage any logistics dependent on the ever-changing data. Typically, in research the best way to address this challenge is to use a tracking system. In this talk, I will discuss how I used R and {shinydashboard} to create an internal clinical tracking dashboard. This ecosystem of R software created a centralized location for lead investigators and project managers to see our data collection progress via interactive data visualizations, provided analysts visibility into the characteristics of our cohort, and acted as an operational tool for research staff that allowed them to independently answer data questions and move their own work forward.  
##   
##   
## Your Score (1-9):   
##   
## name: Robert L. Lobato, MD MS   
## org: Department of Anesthesiology, University of Nebraska Medical Center  
## title: Chief of Multispecialty Anesthesiology   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: I am an anesthesiologist and Chief of the Division of Multispecialty Anesthesiology in the Department of Anesthesiology at the University of Nebraska Medical Center. I have been programming with R for approximately 15 years and develop applications to improve patient care and clinical operations within perioperative medicine.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: R and Shiny Dashboards Facilitate Quality Improvement in Anesthesiology and Perioperative Care  
## abstract: Quality improvement studies quantify progress made toward implementing evidence-based guidelines into clinical care. The Department of Anesthesiology at the University of Nebraska Medical Center developed a dashboard using R and Shiny to track several perioperative quality metrics developed by the Centers for Medicare and Medicaid Services. Despite high performance on a quality measure evaluating rates of postoperative nausea and vomiting (PONV) in high-risk patients, we continued to observe unacceptably high rates of PONV after surgery. A reanalysis of patient-level data identified subsets of patients, not included in the high-risk group addressed in CMS metrics, who were being under-treated according to current perioperative society guidelines. An improvement was made to our Shiny dashboard to place more emphasis on deviations from guidelines-based perioperative care and the resulting rates of complications in those patients. This quality improvement study demonstrates the utility of R and Shiny dashboards in identifying areas where patient care deviates from evidence-based guidelines and illustrates the importance of quality improvement in guiding clinical practice.  
##   
##   
## Your Score (1-9):   
##   
## name: VP Nagraj   
## org: Signature Science  
## title: Bioinformatics Data Scientist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: NA  
## gender: NA  
## poc: NA  
## submission\_type: lightning  
## abstract\_title: twoxtwo: An R Package for Data Analysis with Two-by-Two Tables  
## abstract: The two-by-two contingency table is a commonly used epidemiological method for summarizing relationships between binary exposures and outcomes. The twoxtwo R package (https://github.com/vpnagraj/twoxtwo) provides a dedicated interface for organizing and analyzing observation-level data in two-by-two tables. Functionality includes tools to compute measures of effect (odds ratio, risk ratio, and risk difference), calculate impact numbers and attributable fractions, and perform hypothesis testing. All computed measures yield point estimates as well as intervals with lower and upper bounds corresponding to the specified confidence level. While the package introduces a new S3 class for twoxtwo objects, outputs of analysis functions maintain a tidy structure. With its targeted collection of categorical data utilities, twoxtwo aims to provide a convenient solution for epidemiological investigation of binary exposures and outcomes using R.  
##   
##   
## Your Score (1-9):   
##   
## name: Layla Bouzoubaa   
## org: University of Miami Miller School of Medicine  
## title: Lead Research Analyst   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Layla Bouzoubaa is a Lead Research Analyst for the University of Miami (UM) Miller School of Medicine's Department of Public Health Sciences (DPHS) where she focuses on substance use research. Her work primarily studies patients who suffer from opioid addiction - which includes developing Shiny apps for resource mapping, data harmonization, population statistics/visualizations and modeling that will identify usage patterns for better treatment. Layla also has the pleasure in co-instructing "Data Science for Biostatistics", a 6-week summer course for masters and PhD level students within DPHS.  
##   
## Outside of work, Layla organizes R-Ladies Miami, an international organization that promotes gender equity among the data science community, and Google Developers Group (GDG) Miami, a community group that encourages and enables the discussion of new technologies in South Florida. She also organizes the annual Women in Data Science (WiDS) Miami conference, is a Google Women Techmaker (WTM) ambassador, and is a facilitator of Google's #IamRemarkable initiative.  
## gender: female  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: DOPE: An R Package for Processing & Classifying Drug Names  
## abstract: DOPE (Drug Ontology Parsing Engine) is an R package that aims to synthesize and harmonize drug names from multiple sources, allowing researchers to look up drugs by "brand", "generic" or "street" names.  
## DOPE has built a lexicon with drug names grouped into 11 categories (i.e.,  
## stimulants, depressants, designer drugs, etc.) and 41 classes (e.g., amphetamines, cocaine, etc.) based on categorization provided by the DEA and synonyms/slang provided on publicly available websites.   
## It contains the datasets and a function needed to parse a set of potential drug names and classify them by their class and type. The functions automatically classify any sets of words, which can be a mixture of slang or formal names. A summary function provides summaries when a drug type/class is ambiguous or if a substance is a combination of multiple drugs (i.e., speedball).DOPE can be the backbone for future data science projects, allowing people to efficiently collect  
## and synthesize drug data while dealing with ambiguous terms. For example, DOPE can be used during data collection to prompt clarification for uncommon slang that the DEA has identified, such as "cheese", which can refer to heroin, marijuana or methamphetamine.  
##   
##   
## Your Score (1-9):   
##   
## name: Ethan Heinzen   
## org: Mayo Clinic  
## title: Senior Data Science Analyst   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: NA  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: An Arsenal of R Functions for Statistical Summaries  
## abstract: arsenal is an R package designed to make statistical summaries easy. It includes many functions which the user will find useful to have in his/her "arsenal" of functions. There are 6 main functions, each of which is motivated by a SAS procedure or local macro of similar functionality.  
## Its reporting capabilities are streamlined to work within the latest reporting tools in R and RStudio and use formulas and versatile summary statistics for tables and models. The primary functions include tableby(), a Table-1-like summary of multiple variable types 'by' the levels of one or more categorical variables; paired(), a Table-1-like summary of multiple variable types paired across two time points; modelsum(), which performs simple model fits on one or more endpoints for many variables (univariate or adjusted for covariates); freqlist(), a powerful frequency table across many categorical variables; comparedf(), a function for comparing data.frames; and write2(), a function to output tables to a document.  
##   
##   
## Your Score (1-9):   
##   
## name: Serdar Balci   
## org: Memorial Hospital Group  
## title: Pathologist   
## time zone: (GMT+02:00) Athens, Istanbul, Minsk  
## bio: NA  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Generating reports with R for anatomic pathology laboratory quality control  
## abstract: Anatomic pathology reports are semi-structured texts, with specific jargon that include sensitive patient information. Laboratory information systems provide statistics and quality control measures but more information is needed from a clinical view. R is a practical way to get more insight from pathology reports.  
##   
## Istanbul Memorial Healthcare Pathology, serves 8 hospitals in 5 cities. The hospitals use data systems from different vendors and extracted data are available in various formats (csv, json, excel and pdf), but which can be joined via unique patient and biopsy number.  
##   
## Reports are generated via bookdown. Data are pre-processed in the first chapters, saved as separate RDS files, then read in other chapters as necessary. Bookdown gives flexibility to add new analysis in any place. The reports are rendered with cron jobs.  
##   
## Various packages and regular expressions are used to label pathology reports, follow up of patients, and generate correlations between cytology-pathology and repeat biopsies. Specimen movements, transfer-reporting time measurements, and laboratory physician workload are calculated. Per disease patient survival analyses are also evaluated.  
##   
##   
## Your Score (1-9):   
##   
## name: Daniel T Holmes   
## org: University of British Columbia  
## title: Medical Biochemist   
## time zone: (GMT-09:00) Alaska  
## bio: Daniel Holmes earned his undergraduate degree in Chemical Physics from the University of Toronto. He went to medical school at the University of British Columbia (UBC) where he also did his residency in Medical Biochemistry. He is a Clinical Professor of Pathology and Laboratory Medicine at UBC and Head and Medical Director of the Department of Pathology and Laboratory Medicine at St. Paul’s Hospital in Vancouver. Interests include clinical endocrinology with a focus on secondary hypertension, lipidology, clinical mass spectrometry and laboratory medicine informatics in application to data automation, visualization and clinical utilization.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: R and RShiny for PCR Instrument Communication, Pooling Workflow and Getting Rid of those Nasty Transcriptions  
## abstract: The COVID response of our laboratory mandated rapid upscaling of PCR resulting using all available instruments. Reagent shortages for our fully automated PCR system meant we had to rely on more manual 96-well nucleic acid extractors and thermocyclers. These devices do not have inter-instrument communication and so sample IDs and results were being moved between devices using USB sticks and manual transcription. Further, such a workflow does not permit data integrity required for sample pooling. We developed RShiny applications to track samples from the liquid handler to the extractor, thermocycler and laboratory information system in a 6 week timeframe. This software also managed a 4/6/8/10-plex sample pooling workflow and connected 10 devices and two Linux servers. Ease of development has now permitted all lower volume virology tests to flow through the app avoiding 170,000 manual transcriptions to date. Integration of a Random Forest model has permitted identification of samples more likely to be positive based on patient demographic and geographic information and prior testing. These samples can be diverted from the pooling workflow by scanning at the time of racking, decreasing the functional positivity rate by about 50%.  
##   
##   
## Your Score (1-9):   
##   
## name: Sanjeev Sariya   
## org: Columbia University Medical Center  
## title: Staff associate   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I am a bioinformatics analyst at the Columbia University Medical Center (NY, US). My research focuses on admixed population in the context of Alzheimer's Disease using cohorts with large sample sizes where I am primarily involved in developing and analysis pipelines employing linear mixed models methods. I did my Masters and Bachelors in Bioinformatics from the US and India respectively.   
## Outside of research I take interests in reading (fiction, non fiction), cycling.  
## gender: male  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: GARCOM: A user-friendly R package for genetic mutation counts  
## abstract: Next-generation sequencing (NGS) has enabled analysis of rare and uncommon variants in large study cohorts. A common strategy to overcome these low frequencies and/or small effect sizes relies on collapsing strategies, i.e. to bin variants within genes/regions. Several tools are now available for advanced statistical analyses however, tools to perform basic tasks such as obtaining allelic counts within defined genetics boundaries are unavailable or require complex coding. GARCOM library, an open-source freely available package in R language, returns a matrix with allelic counts within defined genetic boundaries. GARCOM accepts input data in PLINK or VCF formats, with additional options to subset data for refined analyses.  
##   
##   
## Your Score (1-9):   
##   
## name: Deeksha S Bidare   
## org: Baylor College of Medicine  
## title: Medical Student   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: I am a rising third year medical student at the Baylor College of Medicine. I am interested in pursuing a career in general surgery or one of its subspecialties, specifically vascular surgery or trauma/critical care. My research thus far has focused on racial and socioeconomic disparities in treatment of vascular disease processes (carotid stenosis, abdominal aortic aneurysms, and peripheral artery disease). I have used and continue to use R throughout all of my projects for wrangling of big patient datasets, data analysis and interpretation, and data visualization.  
## gender: female  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: Implementation of Big Data Analytics in Exploring the Effect of Varying Health Insurance on Peripheral Artery Disease Outcomes  
## abstract: Medicaid coverage among patients with peripheral artery disease (PAD) is associated with higher rates of leg amputation and worse outcomes. The goal of our analysis was to identify differences and associations with amputation rates among PAD patients with varying health insurance that may explain the outcome disparities  
##   
## De-identified patient-level data on Texas hospitalizations from 2010 to 2014 was collected. Dplyr and tidyverse R packages were used in combination with ICD-9 codes to select for patients who met the inclusion criteria: those who underwent treatment for PAD related foot complications. Several R base functions were used to categorize patients by insurance coverage. Functions from the R stats package were used to conduct variance and regression analyses on several patient, hospital, and geographic factors to identify any differences among the insurance groups and associations of any factors with amputation rates.   
##   
## Higher leg amputations among Medicaid patients appear to be influenced by race, disease severity, hospital characteristics, and outpatient evaluation rates, but not with provider density and location. Differences, associations, and geographic trends were visualized using the R ggplot2 and tigris packages.  
##   
##   
## Your Score (1-9):   
##   
## name: Francisco Requena   
## org: Imagine Institute  
## title: PhD Student   
## time zone: (GMT+01:00) Brussels, Copenhagen, Madrid, Paris  
## bio: Francisco Requena is a PhD Student in the Clinical Bioinformatics lab at the Imagine Institut, Paris. He develops bioinformatics tools for the clinical interpretation of copy number variants (CNVs) in patients with rare diseases.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Building an interpretable machine learning model API for the clinical interpretation of CNVs in patients with rare diseases  
## abstract: Copy number variants (CNVs) play an important role in many rare diseases. Despite their relevance, the clinical interpretation of CNVs is challenging and time-consuming. Moreover, current NGS technologies report a median of 7,439 structural variants per genome, making tool development more urgent.  
##   
## To address this problem, we developed CNVscore, a machine learning (ML) approach designed to aid in the clinical interpretation of CNVs. To build CNVscore, we used the R package rtemis to train a rule-based model (RuleFit) and rstanarm to create a Bayesian model. We combined these two approaches (rule-based and Bayesian) to create an interpretable model capable of quantifying the uncertainty level for each prediction. This approach achieved performance comparable to popular ML models (e.g. XGboost, random forest) while allowing the clinician to know which rules were used to make the prediction and the degree of confidence the model has in that prediction.  
##   
## To make CNVscore easily accessible, we deployed it as an API with the package plumber. In addition, the API will be integrated into our recently published Shiny tool CNVxplorer (http://cnvxplorer.com) so that the user can evaluate the model predictions with complementary information.  
##   
##   
## Your Score (1-9):   
##   
## name: Deniz Topcu   
## org: Baskent University  
## title: Medical Doctor / Biochemistry specialist   
## time zone: (GMT+02:00) Athens, Istanbul, Minsk  
## bio: As a biochemistry specialist, I have a strong background in preanalytical systems, immunoassays, and capillary zone electrophoresis. In addition to my medical education, I also pose a BSc in computer engineering. I have been using R as a primary tool for 4 years. I am using R for day-to-day activities of the clinical lab(e.g. statistics, method verification, machine learning, analytics, etc). I have attended two useR meetings and had a lightning talk.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: How to get insight about your laboratory?  
## abstract: A well-organized laboratory has always a key aspect of informed clinical decisions. The modern laboratory produces an unwieldy amount of data. Data analytic tools can be employed to gain insight from otherwise indecipherable amounts of data for decreasing costs, improving process quality, and increasing patient benefit. Laboratory information systems and middleware can provide insights into the laboratory workflow. However, it is also possible to use R to obtain detailed analytics. In this study developing a web-based analytical tool was aimed at day-to-day usage. Turnaround time, device utilization, staff efficiency, shift workload, test ordering patterns, unnecessary or rarely used tube analysis, and redundant testing analysis can be performed using this tool. Required data sets can be extracted from LIS or middleware (as a spreadsheet or CSV file) and uploaded into the app. Mandatory and optional fields are provided by template files. Uploaded data first checked for consistency and then the analysis will be performed. Analysis results can be followed as numerical or interactive plots. Shiny was used to develop a user-friendly interface. ggplot2 and plotly were utilized for interactive plots.  
##   
##   
## Your Score (1-9):   
##   
## name: Jeffrey Nicholas Fisk   
## org: Yale University  
## title: PhD Candidate   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Originally from New Mexico, Fisk is a RIT alum and current PhD Candidate at Yale University in Computational Biology and Bioinformatics. A graduate of Yale's Cancer Biology Training Program, Fisk's research focuses on the evolution of cancer--particularly in response to therapeutic intervention. Fisk is also interested in education, having performed discipline-based education research, designing and leading teaching pedagogy workshops as a McDougal Fellow, and serving on the Provost's Advisory Committee for Accessibility Resources. In their free time, Fisk enjoys wresting, writing, and playing Pokémon.  
## gender: other  
## poc: no  
## submission\_type: regular  
## abstract\_title: Calculating Cancer Effect Size to Rank Cancer Drivers Contextually with cancereffectsizeR  
## abstract: In cancer research, prevalence of particular genetic alterations is often taken as a proxy for importance. However, prevalence and p-values are not themselves measures of effect. As progression of cancer including the acquisition of therapeutic resistance and the mortality-inducing metastatic spread of therapy-resistant cell populations, is fundamentally an evolutionary process, calculation of selection on these variants is an appropriate effect size metric. Here, we present the current version of cancereffectsizeR, an R package that allows for the calculation of the cancer effect size of somatic variants. In addition to calculating the effect from normal somatic tissue to primary progression, cancereffectsizeR also includes the ability to detect stage-specific selection, epistatic interactions, and therapy-induced selection. Overall, cancereffectsizeR provides a quick, easy way to rank-order variants by effect size from tumor sequencing data, empowering guidance of treatment approaches and discovery of resistance mechanisms.  
##   
##   
## Your Score (1-9):   
##   
## name: Gergely Daroczi   
## org: Rx Studio Inc.  
## title: CTO   
## time zone: (GMT+01:00) Belgrade, Bratislava, Budapest, Ljubljana, Prague  
## bio: Gergely Daroczi is an enthusiast R user and package developer, Ph.D. in Sociology, former assistant professor, currently working in industry with 15 years of experience in data science, engineering, cloud infrastructure and data operations at SaaS, fintech and adtech companies, having a strong interest in building a scalable data platforms on the top of R, AWS and other applications. He maintains a dozen of CRAN packages related to using R in production (automated reports, logging, database connections, API integrations), co-authored a number of journal articles in social and medical sciences, and wrote a book on "Mastering Data Analysis with R".  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Democratizing Access to Personalized Medicine Using R  
## abstract: Model-Informed Precision Dosing and PK/PD modeling are still not accessible to the masses -- as gated by expensive software suites and/or complicated interfaces requiring special knowledge. This means that most patients are still treated with potentially suboptimal label doses, while personalized medicine could save not only drug and related costs, but lives as well with better health outcomes.  
##   
## Rx Studio is an R-based cloud platform providing clinical decision support for precision dosing, using peer-reviewed or custom-built PK/PD/PGx models -- heavily relying on open-source tools, and dedicated to democratizing access to pharmacological models through easy access, free for individuals and education and globally inclusive pricing for clinics and health systems.  
##   
## This talk provides an overview on the technical platform, the user-facing interfaces and APIs, using R in a cloud infrastructure, difficulties with compliance and international laws when using open-source software, internationalization, and localization of the web interfaces and R packages as well; then discussing opportunities for partnerships with clinical researches, clinical pharmacists and educators.  
##   
##   
## Your Score (1-9):   
##   
## name: Elizabeth Chun   
## org: Texas A&M University  
## title: Undergraduate Researcher   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Elizabeth Chun is an Undergraduate Researcher working under Dr. Irina Gaynanova in the Texas A&M Department of Statistics. As a current Molecular and Cell Biology major, Elizabeth greatly values bridging the world of biomedical data with statistics and data analysis to further both research and clinical practice. In her free time, Elizabeth enjoys playing piano and exploring nature.  
## gender: female  
## poc: NA  
## submission\_type: regular  
## abstract\_title: Interpreting blood glucose data with R package iglu  
## abstract: Continuous Glucose Monitoring (CGM) data play an increasing role in clinical practice as they provide detailed quantification of blood glucose levels during the entire 24-hour period. The R package iglu implements a wide range of CGM-derived metrics for measuring glucose control and glucose variability. The package also allows one to visualize CGM data using time-series and lasagna plots. A distinct advantage of iglu is that it comes with a point-and-click graphical user interface (GUI) which makes the package widely accessible to users regardless of their programming experience. Thus, the open-source and easy to use iglu package will help advance CGM research and CGM data analyses. R package iglu is publicly available on CRAN and at https://github.com/irinagain/iglu.  
##   
##   
## Your Score (1-9):   
##   
## name: Sil   
## org: Maastricht University  
## title: Assistant professor   
## time zone: (GMT+01:00) Brussels, Copenhagen, Madrid, Paris  
## bio: Dr. Sil Aarts works as an assistant professor at the Department of Health Services Research, part of the Care and Public Health Research Institute (Caphri), Maastricht University. She is affiliated with Living Lab in Ageing and Long-Term Care, a structural partnership between nine long-term care (LTC) organizations, a university of applied sciences, Gilde Intermediate Vocational Training Institute, VISTA College (secondary vocational education)  
## and the university. Her line of research focuses on the use of (big) data in LTC in order to improve quality of care for clients, quality of life for clients and their loved ones, and the quality of work for healthcare professionals.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: Text mining in long-term care: exploring the usefulness of computer-aided methods for investigating quality of care  
## abstract: Narrative data are often collected to evaluate quality of care. This results in large amounts of textual data. As the volume of data increases, it becomes beyond the capability of humans to analyse it. This study aims to explore the usefulness of text-mining regarding narrative data on quality of care. Data was collected as part of ‘Connecting Conversations’: assessing experienced quality of care by interviews with residents of nursing homes (n=39), family members (n=37) and care professionals (n=49). Word frequencies, a correlation analysis and a sentiment analysis were conducted. A survey was conducted to establish a sentiment analysis model. Residents, family members and care professionals uttered respectively 285, 362 and 549 words per interview. Words that occurred most frequently in the interviews are often positive. Although there are some differences in wording, correlation analysis displayed that similar words are used by all groups to describe quality of care. Care professionals are more diverse in their sentiment than residents and family members: while some express a positive sentiment, others express more negativity. This study demonstrates the usefulness of a text-mining to extend our knowledge on quality of care.  
##   
##   
## Your Score (1-9):   
##   
## name: Sophia Shalhout   
## org: Massachusetts General Hospital, Harvard Medical School  
## title: Cutaneous Oncology Research Fellow   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Sophia Shalhout, PhD joined the Miller Lab at MGH/Harvard Medical School as the Cutaneous Oncology Research Fellow in 2019. Dr. Shalhout is currently using her skills in Biology and Data Science to identify biomarkers of response to therapies for advanced/rare skin cancers and developing clinical informatics tools for the multi-institutional Merkel Cell Carcinoma Patient Registry.  
## gender: female  
## poc: yes  
## submission\_type: regular  
## abstract\_title: eLAB: A Large-Scale Laboratory Integration Informatics Pipeline for Clinical Research  
## abstract: Objective: To develop a clinical informatics pipeline designed to capture large-scale structured electronic health record (EHR) data for a national patient registry.   
## Materials and Methods: The EHR-R-REDCap pipeline is implemented using R-statistical software to remap and import structured EHR data into the REDCap-based multi-institutional Merkel Cell Carcinoma (MCC) Patient Registry using an adaptable data dictionary.   
## Results: Clinical laboratory data were extracted from EPIC Clarity across several participating institutions. Labs were transformed, remapped and imported into the MCC registry using the EHR labs abstraction (eLAB) pipeline. Forty-nine clinical tests encompassing 482,450 results were imported into the registry for 1,109 enrolled MCC patients. Data-quality assessment revealed highly accurate, valid labs. Univariate modeling was performed for labs at baseline on overall survival (N=176) using this clinical informatics pipeline.   
## Conclusion: We demonstrate feasibility of the facile eLAB workflow. EHR data is successfully transformed, and bulk-loaded/imported into a REDCap-based national registry to execute real-world data analysis and interoperability.  
##   
##   
## Your Score (1-9):   
##   
## name: Wes Wilson   
## org: University of Pennsylvania  
## title: Cancer Researcher / Bioinformatics   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Wes is a Canadian cancer researcher and tumor immunologist currently conducting clinical trials at the University of Pennsylvania in multiple myeloma using combination CAR-T therapy. They been a R user since 2004.  
## gender: other  
## poc: no  
## submission\_type: regular  
## abstract\_title: The Bachelorette Taught Me How to Give My Final Rose to the Best Integration Method to Remove Batch Effects from My Patient Data  
## abstract: Batch effects plague large patient data studies which can add bias (age, ethnicity, sex) to the results and impact reproducibility when that same data is analyzed by different researchers. This problem is confounded in wet lab data acquisition by adding variables such as date, personnel handling the sample, and technology variations. Different integration methods can help alleviate that issue. But with so many different methods how do you pick which to spend the rest of your analysis with? What if you pick the wrong one and you find out months later, do you settle for good enough or start your analysis from scratch? This can be painful and confusing time in clinical data scientists life. During the pandemic lock down I binge-watch watched a ton of cheesy TV and found myself wondering why the bachelorette would slowly remove the people she thought were not right, only to pick someone in the finale and months later break up and get back together with someone from episode 2. As I tried to work out ways for her to avoid heart ache I thought about applying the same metrics in personally compatibility to my own data to find the best way to solve my own "batch" problems.  
##   
##   
## Your Score (1-9):   
##   
## name: Luke McGuinness   
## org: University of Bristol  
## title: Senior Research Associate   
## time zone: (GMT) Greenwich Mean Time: Dublin, Edinburgh, Lisbon, London  
## bio: Luke McGuinness is an NIHR Doctoral Research Fellow at the University of Bristol Medical School, where he uses large electronic health record (EHR) databases and evidence synthesis methods to examine the relationship between blood lipid levels and dementia risk.  
##   
## As an enthusiastic R user/developer, he maintains several research-orientated open-source R packages including robvis and medrxivr, in addition to some smaller utility/learning packages (e.g. pathformatr).  
##   
## Outside of work, he is an avid fan of running, LEGO and really terrible movies.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Incorporating risk-of-bias assessments into evidence syntheses with robvis  
## abstract: Risk-of-bias assessment is regularly hailed as an essential part of the evidence synthesis process, providing a structured way to critically assess the internal validity of included studies. However, evidence from meta-research studies has demonstrated that while risk-of-bias assessments are performed, the data produced by the assessments are infrequently visualised or incorporated into statistical analyses.   
##   
## {robvis} is an established R package which aims to solve this issue by making it easier for reseachers to create reproducible publication-quality risk-of-bias plots in line with best-practice guidelines. This presentation for R Medicine will introduce the robvis package, outline the motivation for its development, and showcase how users can use the package to report risk-of-bias assessments in their own reviews. It will also introduce new functionality in robvis, focused on better integration with the {metafor} package for performing meta-analysis, which allows users both to present statistical analyses and risk-of-bias results side-by-side (via graphics known as "paired forest plots") and to readily stratify their analysis by the level of risk of bias in each included study as part of a sensitivity analysis.  
##   
##   
## Your Score (1-9):   
##   
## name: Christian Röver   
## org: University Medical Center Göttingen  
## title: Research Associate   
## time zone: (GMT+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna  
## bio: Christian Röver is a research associate at the Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany. After studying Statistics at Dortmund University and Iowa State University, he earned a PhD degree at The University of Auckland. While his masters thesis was on classification methods, the PhD thesis was on computer intensive methods for Bayesian parameter estimation. After his PhD, he went on to work mostly on signal detection and parameter estimation problems at the Max Planck Institute for Gravitational Physics (Albert Einstein Intitute) in Hannover, before moving to medical statistics at the University Medical Center Göttingen. His currently methodological research focus is on Bayesian methods for meta-analysis and their implementation in R.  
## gender: NA  
## poc: NA  
## submission\_type: regular  
## abstract\_title: Bayesian random-effects meta-analysis using bayesmeta  
## abstract: Meta analyses are useful in many medical fields. A Bayesian approach to inference is very attractive in this context, especially when a meta-analysis is based only on few studies. The bayesmeta R package provides readily accessible tools to perform Bayesian meta-analyses and generate plots and summaries, without having to worry about computational details. It allows for flexible prior specification and instant access to the resulting posterior distributions, including prediction and shrinkage estimation, and facilitating for example quick sensitivity checks. This presentation will showcase its use and also provide pointers to more advanced applications.  
##   
##   
## Your Score (1-9):   
##   
## name: Sean Meyer   
## org: Michigan Medicine, University of Michigan  
## title: PhD Candidate   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I'm a PhD candidate in the Design Science Program at the College of Engineering. My work focuses on optimizing the development of early warning system models through the identification of common design elements present in the data preparation process. These elements can then be used as a framework for turning clinically relevant concepts into modeling-ready structured data.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: gpmodels: A Grammar of Prediction Models  
## abstract: The gpmodels R package provides a grammar for data preparation and evaluation of fixed-origin and rolling-origin prediction models using data collected at irregular intervals (e.g., electronic health record data). Since predictions often begin and end at specific times for each patient (e.g., hospital admission or discharge), gpmodels provides clean, pipeable functions for the creation of predictors and outcomes that are linked to either a fixed origin (e.g., single observation per patient) or rolling origin (e.g., multiple observations per patient). Creation of machine learning-ready is highly CPU- and memory-intensive, so gpmodels supports parallel processing (using furrr), allows processing individual chunks in parallel, and can write output directly to file to minimize memory usage. gpmodels also provides functions to preprocess data elements (e.g., dummy coding of time-series factor variables) and for the evaluation of time-series models. The model evaluation code in gpmodels is based on the code used in recent publications evaluating performance of the Epic Sepsis Model and the Epic Deterioration Index.  
##   
##   
## Your Score (1-9):   
##   
## name: Jie Cao   
## org: University of Michigan  
## title: PhD Student   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: NA  
## gender: female  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: runway: An R Package to Visualize Prediction Model Performance  
## abstract: Runway is an R package that facilitates the visual evaluation and comparison of prediction models and provides publication-ready graphics. Visualizing performance is an important way to communicate a prediction model’s strengths and limitations. The core visualizations implemented in runway are threshold-performance plots, calibration plots, and receiver operating characteristic (ROC) curves. Threshold-performance plots provide a clean, faceted approach to demonstrating sensitivities, specificities, and positive and negative predictive values across a range of probability thresholds. Calibration curves show either binned (e.g., deciles) or continuous measures of observed risk across a range of predicted risk. Models with different types of miscalibration can be difficult to compare on a threshold-performance plot, so the package also supports traditional ROC curves. The visualizations include 95% confidence intervals, can be used to plot either a single model (black and white) or multiple models (in color), and include distributions of predicted risk (as histograms for single models and density plots for multiple models).  
##   
##   
## Your Score (1-9):   
##   
## name: Ariel Mundo   
## org: University of Arkansas  
## title: Graduate Research Assistant   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Ariel Mundo is a PhD Candidate and former Fulbright fellow in the Department of Biomedical Engineering at the University of Arkansas. His research is focused on detecting tumor response in colorectal cancer using optical and molecular biology techniques. He is an avid R user, and is interested in data visualization, statistical methods for longitudinal data that can be used in biomedical research, and research reproducibility.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Generalized additive models for longitudinal biomedical data  
## abstract: Biomedical research frequently uses longitudinal studies to capture the evolution of an effect over time. Traditionally, such data is analyzed using repeated measures ANOVA (rm-ANOVA) models or linear mixed models (LMEM; but it is frequently the case that those models are not appropriate because the trend of the data is not linear. This non-linear behavior in longitudinal data occurs frequently in biomedical research, but its implications from a Statistical perspective are often not considered, leading to unreliable inference.  
## Generalized additive models (GAMs) are a class of models that allow non-linear trends in the data and that can provide robust inference, but they are little known in the biomedical realm.   
##   
## This talk will provide a brief summary of examples in the biomedical literature where non-linear trends in data have been observed, some of the limitations of linear models that make them unusable in non-linear data, and will show a brief implementation of GAMs using R in simulated data that follows reported trends in the literature.  
##   
##   
## Your Score (1-9):   
##   
## name: David Bock   
## org: Göteborg University  
## title: Statistician   
## time zone: (GMT+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna  
## bio: David Bock is biostatistician and associate professor of epidemiology. Main area is the contribution to improve clinical decision making in the treatment of patients diagnosed with colorectal cancer prostate cancer or perforated diverticulitis.  
## gender: NA  
## poc: NA  
## submission\_type: lightning  
## abstract\_title: Assessing variable importance in clinical prediction models using R  
## abstract: Besides the aim of predicting future values of the outcome we may be interested in evaluating which variables and to which extent they contribute to the predictions. To ensure external validity this must be addressed in a setting where all sources of uncertainty are accounted for. The aim is to give an illustration on how to assess and graphically quantify variable importance in a stepwise approach where LASSO and unpenalized logistic regression is combined and where uncertainty is assessed by boostrapping techniques.  
##   
##   
## Your Score (1-9):   
##   
## name: Victor Castro   
## org: Mass General Brigham  
## title: Data Scientist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Victor is a Data Scientist at Mass General Brigham, a large health care system in the Boston area. He works on machine learning algorithms using electronic health record data, including for clinical risk prediction, phenotyping and population health.  
## gender: male  
## poc: yes  
## submission\_type: regular  
## abstract\_title: Assessing Machine Learning Model Performance in Diverse Populations and Across Time  
## abstract: In recent years numerous machine learning models to predict clinical risk have been reported in the research literature. Many of these models are developed in large and diverse databases of electronic health record or other medical datasets. Few risk models, however, have been evaluated for potential bias among subgroups of patients in the training or evaluation set. In this talk we’ll cover our experience training a model to stratify risk among patients hospitalized for COVID-19 during the early course of the pandemic and then performing subsequent evaluation of the model performance a year later. We’ll describe how we extend the tidymodels yardstick package to evaluate performance across time as well as subgroups of patients. Our results illustrate the importance of investigating risk stratification models across patient subgroups, as a step toward ensuring that particular groups of patients are not adversely impacted by the application of machine learning models in healthcare.  
##   
##   
## Your Score (1-9):   
##   
## name: Beth Atkinson   
## org: Mayo Clinic  
## title: Principal Biostatistician   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Beth Atkinson started working with Splus in 1990 and soon was helping develop the rpart package. Since then she has helped develop other packages including arsenal and kinship2. She has actively been working with Terry Therneau on the survival package for several years and they, along with another colleague, are working on a book on the topic.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: Multistate data using the survival package  
## abstract: Survival data, also known as time-to-event data, is very common in medical research. Historically, analysis focused on single endpoints such as death or cancer recurrence. Increasingly, research questions are more complex and need multistate transition models. This approach allows you to handle situations where subjects move between states, such as healthy to illness to death. The survival package has had functions to handle standard survival analysis for many years. Starting with version 3.2 there are tools to handle multistate models as well. The overall goals for the additions to the packages were to:   
## \* Make probability-in-state curves as easy to create as a Kaplan-Meier curve   
## \* Make multistate models as easy to fit as a Cox model  
## \* Provide access to other estimates of absolute risk, e.g., the restricted mean time in state.  
##   
## This presentation will highlight questions that can be addressed with multistate models and will show how to code analyses to answer the questions.  
##   
##   
## Your Score (1-9):   
##   
## name: Miguel Cosenza   
## org: University Medical Center Freiburg  
## title: PhD researcher   
## time zone: (GMT+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna  
## bio: I am a biologist from Venezuela currently working in the interphase between clinical proteomics and data science as part of my Ph.D. research in Germany. As a master's student in Brazil working with large-scale proteomics of host-parasite interactions, I got interested in data science and reproducible research as a way to extract meaningful information from high-dimensional biological data. Currently, while combining experimental work and data science, I question mass-spectrometry-based proteomics data through the use of supervised and unsupervised statistics to identify proteins associated with particular clinical conditions in the context of cancer and heart disease.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Proteomics of reverse cardiac remodeling: extracting biological meaning from high-dimensional data  
## abstract: In the context of cardiac disease, the limited regenerative capacity of the heart can eventually lead to heart failure (HF). Currently, left ventricular assist device (LVAD) implantation is one of the two available therapies. However, in a subset of patients, LVADs promote cardiac unloading and recovery of heart function, allowing LVAD explantation. Characterizing the molecular mechanisms associated with this so-called reverse cardiac remodeling (RCR) is crucial for the identification of predictive biomarkers of cardiac recovery for their use in personalized medicine schemes. Using a mouse model of RCR and mass spectrometry-based proteomics, we were able to identify and quantify thousands of proteins from left ventricular tissue samples. In this talk, I would showcase the use of R for the analysis of this high dimensional data, using (1) visualization approaches (ggplot2) for quality control of the quantitative data, (2) data wrangling methods to explore intrinsic protein degradation processes, and (3) tools for unsupervised (mixOmics) and supervised (limma) analyses to both explore the effect of RCR on protein abundance and identify potential molecular signatures of cardiac recovery.  
##   
##   
## Your Score (1-9):   
##   
## name: Timothy Keyes   
## org: Stanford University School of Medicine  
## title: MD/PhD student   
## time zone: (GMT-08:00) Pacific Time (US and Canada); Tijuana  
## bio: Timothy is an MD/PhD student studying cancer biology and biomedical informatics at the Stanford University School of Medicine. As a biomedical data scientist, Timothy's research focuses on the application of machine learning to single-cell data analysis in the context of pediatric leukemia. Through the use of emerging, high-throughout single-cell technologies such as mass cytometry and sequence-based cytometry, Timothy’s work centers on building predictive models of patient outcomes - such as relapse or minimal residual disease (MRD) - at the point of diagnosis. To do so, he uses a variety of computational tools including generalized linear models, clustering, and deep learning. In addition, his work prioritizes constructing easy-to-use, highly-reproducible data analysis pipelines that can be shared as open-source tools for the scientific community.  
## gender: other  
## poc: no  
## submission\_type: regular  
## abstract\_title: {tidyTOF}: Building a reproducible pipeline to predict patient outcomes from single-cell data using tidy data principles  
## abstract: In recent years, the concept of “tidiness”—a data analysis framework in which data frames are structured such that each variable is a column and each observation is a row—has become a staple of the R community. This is largely because working with “tidy data” makes it easy to map the meaning of a dataset to a specific and consistent structure, which in turn allows for the development of easily-used data analysis tools for a variety of applications. Despite these strengths, relatively few tools have been developed for analyzing clinical and biomedical data using tidy data principles.  
##   
## Here, we present {tidytof} – a novel R package for tidy analysis of mass cytometry (CyTOF) single-cell data – as a case study for using tidy data principles to simplify the manipulation, visualization, and modeling of biomedical data. Specifically, {tidytof} leverages tidy data analysis to implement a reproducible and human-readable pipeline for common single-cell analysis tasks including reading/writing data, clustering, dimensionality reduction, feature engineering, and patient-outcomes modeling. Our presentation will highlight the process behind and benefits of using tidy principles to build simple and intuitive biomedical data analysis workflows.  
##   
##   
## Your Score (1-9):   
##   
## name: David M. Miller   
## org: Massachusetts General Hospital  
## title: Medical Oncologist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: David M. Miller, MD, PhD is an Instructor in Dermatology and Medicine at Harvard Medical School and member of the Department of Dermatology and the Department of Medicine at Massachusetts General Hospital, where he is the director of the Center for Merkel cell carcinoma (MCC). Dr. Miller received his medical (MD) and PhD degree in Biomedical Sciences from the University of Massachusetts. While at UMass, Dr. Miller received the American Diabetes Association Physician-Scientist Training Award for three consecutive years, a Fuller Foundation grant, and he was inducted into the Alpha Omega Honor Society. He then completed a Dermatology residency at Columbia University Medical Center, followed by a residency in Internal Medicine at Brigham and Women's Hospital. Following both residencies, Dr. Miller completed fellowship training in Hematology/Oncology at Beth Israel Deaconess Medical Center and a fellowship at the Food and Drug Administration.   
## Dr. Miller’s research laboratory conducts clinical and translational research in non-melanoma skin cancer. He is the principal investigator of several clinical trials in MCC and squamous cell carcinoma (SCC), where he is evaluating new treatment approaches in the neoadjuvant, adjuvant and refractory disease setting. In order to complement clinical trial efforts, which have inherent limits in external validity, his lab is developing clinical informatic platforms to optimize real world data collection. In doing so, they have formed a strategic partnership with Project Data Sphere to build a multi-institutional MCC patient registry. Through this registry they aim to not only better define the natural history and identify best practices in MCC, but also to establish a framework to collect and analyze data in rare diseases. As co-chair of the Project Data Sphere Merkel cell carcinoma registry Task Force, Dr. Miller has been the chief architect of the data collection and data analytics effort in MCC. His lab has built a multi-dimensional data collection tool that captures clinical and genomic information across the patient journey. Their platform incorporates automated data flows from the electronic health record to augment classical data abstraction. In addition, they are developing data science tools to facilitate real world data analysis. In an effort to disseminate their insights on real world data collection and analysis, they are both publishing them in peer-reviewed journals and making them freely available on our lab website (themillerlab.io). These research efforts have been supported by a Loan Repayment Grant funded by the NIH, a Research Grant sponsored by the American Skin Association, the ECOG-ACRIN Cancer Research Group’s Paul Carbone, MD Fellowship Award, and the Project Data Sphere’s Mentoring Award in Oncology, Advanced Analytics and Regulatory Science.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: GENETEX - A GENomics Report TEXt Mining R Package and Shiny Application Designed to Capture Real-World Clinico-Genomic Data  
## abstract: Objectives: Clinico-Genomic Data (CGD) acquired through routine clinical practice has the potential to improve our understanding of clinical oncology. However, these data often reside in heterogeneous and semi-structured data, resulting in prolonged time-to-analyses.  
## Materials and Methods: We created GENETEX: an R package and Shiny application for text mining genomic reports from EHR and direct import into REDCap.  
## Results: GENETEX facilitates the abstraction of CGD from EHR and streamlines capture of structured data into REDCap. Its functions include natural language processing of key genomic information, transformation of semi-structured data into structured data and importation into REDCap. When evaluated with manual abstraction, GENETEX had >99% agreement and captured CGD in approximately one-fifth the time.  
## Conclusions: GENETEX is freely available under the Massachusetts Institute of Technology license and can be obtained from GitHub. GENETEX is executed in R and deployed as a Shiny application for non-R users. It produces high-fidelity abstraction of CGD in a fraction of the time.  
##   
##   
## Your Score (1-9):   
##   
## name: Marcus Adams   
## org: Merck  
## title: Associate Director, Engineering   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Marcus Adams earned his BEng and MS in Chemical Engineering from the University of Delaware and Villanova University, respectively. His more than decade at Merck spans the bio-pharmaceutical spectrum and includes experience in pre-clinical PK/PD modeling, product commercialization, in-line technology support, procurement, and vaccine distribution technology development. Currently he works as a part of the End to End Technical Operations Digital and Data CoE team, leveraging Merck’s Big Data Platform in the development of manufacturing information data models, report automation tools, and integrated-systems analysis applications. His professional interests include effective digital visualization, reproducible research/analysis, and convincing his coworkers of the diverse, flourishing world beyond Microsoft Excel.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Second Server to the Right and Straight On 'til Production: Deploying a GxP Shiny Application  
## abstract: In the classic software development text “The Mythical Man-Month” Fred Brooks describes a 9x difference in the effort between creating a program (think a simple app that you run only on your own personal computer) and a programming systems product (or what we may call an app in production). According to a 2019 Burch Works study, however, ~80% of data scientists don’t come from a formal computer science background and therefore likely haven’t read this book. More importantly, though, this also means that they likely do not have experience putting an application or model into production use. They can create wonderous neural networks for splendorous NLP use-cases but lack the experience addressing all the other the requirements which create a robust and dependable product for end users.   
##   
## R can and has been put into production by many companies, both large and small. This talk will share insights from the Merck Manufacturing Digital and Data CoE’s experience in launching a GxP reporting Shiny application into production. The speaker, who comes from a chemical engineering background, will share what he learned about the other 89% of what it takes to deploy reliable, secure, and maintainable production applications.  
##   
##   
## Your Score (1-9):   
##   
## name: Will Landau   
## org: Eli Lilly and Company  
## title: Research Scientist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Will Landau is a Research Scientist at Eli Lilly and Company, where he develops methods and tools for clinical trial simulation and analysis, and he is the creator and maintainer of the targets and stantargets R packages. Will earned his PhD in Statistics at Iowa State University in 2016, where his dissertation research applied Bayesian methods, hierarchical models, and GPU computing to the analysis of RNA-seq data.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Target Markdown and stantargets for Bayesian model validation pipelines  
## abstract: The {targets} R package enhances the reproducibility, scale, and maintainability of data science projects in computationally intense fields such as machine learning, Bayesian Statistics, and statistical genomics. Recent breakthroughs in the {targets} ecosystem make it easy to create ambitious, domain-specific, reproducible data analysis pipelines. Two highlights include Target Markdown, an R Markdown interface to transparently communicate the entire process of pipeline prototyping and construction, and {stantargets}, a new rOpenSci package that generates specialized workflows for Stan models while reducing the required volume of user-side R code. This presentation demonstrates both capabilities in a simulation-based workflow to validate a Bayesian longitudinal linear model common to clinical trial data analysis.  
##   
##   
## Your Score (1-9):   
##   
## name: Piru Perampalam   
## org: Procogia  
## title: Senior Bioinformatics Engineering Consultant   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I did my PhD in Biochemistry at Western University and worked on several cancer research projects at London Health Sciences Centre in London, Canada. My research focussed on understanding the biology of high-grade serous ovarian cancer (HGSOC) progression and dormancy. I utilized many high-throughput assays including RNA-sequencing and CRISPR screens to identify novel molecular pathways that can be targeted to treat late-stage HGSOC. This called for the development of new experimental methods and computational tools in both R and Python. I am currently at Procogia as a Senior Bioinformatics Engineering Consulting working on exciting bioinformatics projects within the pharmaceutical industry.  
## gender: male  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: R shiny as a collaborative tool: unlock the power of your data  
## abstract: As clinicians and scientists, we are often generating large volumes of data, and more than ever before, collaboration has become a vital component of research over the past year. Unfortunately, the way most of us share and collaborate with our data hasn't evolved; we are still sending large Excel files back and forth. This stifles creativity and progress because the person who created the Excel files is often tasked with changing parameters and re-sharing it with everyone - sometimes a labour intensive process.   
##   
## In this short talk, I will discuss how R shiny can be used make this process more efficient. We will look at a real-world scenario where the power of data manipulation is put into the hands of the entire team - not just one person. We will look at how it improved creativity and encouraged more constructive discussions, ultimately leading to more impactful results.  
##   
##   
## Your Score (1-9):   
##   
## name: Richard Steven Hanna   
## org: Children's Hospital of Philadelphia  
## title: Data Scientist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: As a Data Scientist at Children’s Hospital of Philadelphia, I develop data-driven solutions and design operational software to support clinical needs. The contextual scope of my work has ranged from pediatric cardiac arrest to airway management and, most recently, to cell and gene therapy and stem cell transplantation. I strive to live every day with the ambition and excitement that what I know today is nothing compared to what I will learn tomorrow.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Stem Cell Transplant Outcomes Reporting using R/Shiny  
## abstract: Stem cell transplantation is a life-saving therapy for various malignant and non-malignant diseases. To aggregate real-world data on transplant patients, all cancer centers in the U.S. are required by law to report outcomes such as the time from transplant to the emergence of certain blood cells, known as “engraftment.”  
##   
## Traditionally, engraftment is reported by data managers who manually look up information in medical records and complete web-based forms. This tedious manual process requires integration of multiple data sources and, at a top-tier pediatric cancer center, was found to have error rates as high as 25%.  
##   
## To mitigate this problem, we developed an R/Shiny application that calculates engraftment dates by querying and integrating multiple clinical data sources. We used the {golem} framework and deployed the app on RStudio Connect. The app passed clinical validation and is now in day-to-day use by our data managers, having eliminated the manual process.  
##   
## In this lightning talk, we will outline the clinical validation, production engineering, and operationalization of the engraftment app. We will also discuss how the app could be implemented at other cancer centers.  
##   
##   
## Your Score (1-9):   
##   
## name: Haocheng Wang   
## org: Dana Farber Cancer Institute  
## title: Data Visualization Specialist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Haocheng Wang graduated from St. Olaf College in 2021 with a B.A. in Biology, Mathematics, and Computational Neuroscience as well as concentrations in statistics and mathematical biology. He plans to pursue a Ph.D. in computational biology or biostatistics. Wang has been working as a data visualization specialist in the Murakami lab at Dana-Farber Cancer Institute since the summer of 2020. He is interested in computational modeling, statistical analysis, and data management.  
## gender: male  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: Leukemia Analysis & Data Exploration Repository (LeADER)  
## abstract: Leukemia Analysis & Data Exploration Repository (LeADER) is a data management and visualization platform for cancer clinical trials and correlative laboratory studies at Harvard Medical School. Multidisciplinary research groups face operational inefficiencies and data sharing bottlenecks when performing complex integrative analyses. LeADER bridges the data accessibility gap separating clinicians, bench scientists, bioinformaticians, and statisticians, thereby advancing translational medicine. By linking de-identified clinical, biological, and molecular features of patients, their tumor samples, patient-derived tumor models, and individual tumor cells, collaborators may independently interrogate harmonized, validated, and real-time data.  
## LeADER comprises a network of linked R Shiny apps, each drawing from a common aggregate data store but deploying unique subsets of data with customized visualization and analysis tools for trial-specific user groups. Distinct features build on existing Shiny functionality to facilitate seamless interaction with complex hierarchical data, such as time series and biospecimen provenance with one to many relationships. Since launch, LeADER has yielded efficiency gains in all steps of the data pipeline.  
##   
##   
## Your Score (1-9):   
##   
## name: Paul Schneider   
## org: University of Sheffield  
## title: PhD Student   
## time zone: (GMT+01:00) Brussels, Copenhagen, Madrid, Paris  
## bio: Paul is a PhD student at the Wellcome Trust Doctoral Training Program for Public Health, Economics, and Decision Science at the University of Sheffield. His background is in clinical medicine and health sciences: he is a physician by training, and completed a doctoral degree in medicine at the Institute for Health Systems Research, University of Witten/Herdecke, Germany. He also holds a research masters degree in health sciences from Maastricht University, the Netherlands. He has worked on research projects in a range of different areas, including health technology assessment, digital disease surveillance, health service research, clinical drug trials, and global health policy.  
## gender: NA  
## poc: NA  
## submission\_type: lightning  
## abstract\_title: Shiny Decision Modelling  
## abstract: Statistical models are used ubiquitously across health care to inform decision making. One notable area is health economics, where decision models are a mandatory part of regulatory assessment of the costs and benefits of new treatments.  
##   
## Traditionally, these models have often been built in MS Excel, but as model complexity increases, R is becoming increasingly popular. However, R lacks a simple point and click user interface. This might make the switch from Excel to R seem daunting, and it makes it difficult to directly communicate model results with decisions makers and other stakeholders.  
##   
## R Shiny has the potential to resolve this limitation. It allows developers to embed decision models (e.g. economic models, patient decision aids) developed in R into interactive web-interfaces. Users can specify their own preferences and assumptions about model parameters and run different scenario analyses.  
##   
## In this talk, we provide a brief tutorial on how to wrap a decision model built in R into a Shiny application. As a case study, we use a simple cost-effectiveness model to demonstrate main principles and basic functionality and then explain how this approach can be applied on almost any decision model.  
##   
##   
## Your Score (1-9):   
##   
## name: Kelsey Chalmers   
## org: Lown Institute  
## title: Data Scientist   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Dr Chalmers joined the Lown Institute as a Data Scientist early 2020, where she works on a number of projects on overuse including the Lown Institute Hospital Index. Prior to this, she worked as a health policy researcher at the University of Sydney, Australia, where she also received her PhD in public health, with a focus on measuring overuse of medical care. She graduated from the University of Queensland with a degree in Mathematics and Statistics.  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: A Shiny app to improve transparency and collaborations on developing health service measures from claims data  
## abstract: You’re an analyst with access to claims data in a SAS environment. Your collaborators want to investigate a particular procedure or service, and to develop a definition for that service based on diagnostic code exclusion and inclusion criteria. For example, they might suspect that use of this service is ‘low-value care’, and would like to estimate the proportion of claims that fit this definition.   
##   
## Our SAS program uses a classification tree to find the most important diagnosis codes for predicting whether a claim had your selected service or not, and therefore finds the diagnosis codes most relevant to creating such a definition. The SAS program outputs a table with claim counts of these overlapping codes, which is an input to our R Shiny application. This app lets users select their own inclusion and exclusion codes and immediately see the estimated impacts of these decisions on the number of claims fitting this definition. This is an enormously helpful research tool to guide researchers’ decisions when developing these claims-based measures.  
##   
##   
## Your Score (1-9):   
##   
## name: Marly   
## org: Biogen  
## title: Gotti   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: NA  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Performing Risk Assessments of R Package Accuracy using Shiny  
## abstract: In this presentation we introduce the Risk Assessment Shiny application and explain how it can be used to perform a risk assessment of R package accuracy within a validated infrastructure as required within a biopharmaceutical regulatory setting. The Risk Assessment Shiny Application is an interactive web application that serves as an interface to the riskmetric package. Both tools have been developed by the R validation hub, which is a cross-industry initiative funded by the R Consortium. The shiny application portraits the metrics coming from riskmetric, categorized into maintenance, community usage, and testing. The application can be used to record comments on the metrics, which are stored in an underlying database. This makes it possible to stop and start reviews at the reviewer’s convenience. Once a review is complete, a final summary comment can be provided before a final decision is made on the package. In line with the white paper, the decision is an overall risk score: low, medium, or high. The reviewer is then able to generate either an HTML or a DOCX report containing the metrics, comments, and some additional high-level information about the package.  
##   
##   
## Your Score (1-9):   
##   
## name: Laura Wiley   
## org: University of Colorado Anschutz Medical Campus  
## title: Assistant Professor   
## time zone: (GMT-07:00) Mountain Time (US and Canada)  
## bio: Dr. Wiley develops methods for using data from electronic health records for precision medicine discovery and implementation. She is a principal investigator in the Colorado Center for Personalized Medicine and an Assistant Professor in the Division of Biomedical Informatics and Personalized Medicine at the University of Colorado Anschutz Medical Campus. Her published work involves developing computational phenotyping algorithms for use in EHR-linked biobanks. Her current work focuses on clinical evidence generation from electronic health records and informatics implementation, serving as the lead informatician on an NIH Cancer Moonshot-funded project to create a comprehensive tobacco cessation service at the University of Colorado Cancer Center. Dr. Wiley is actively involved in the informatics community, currently serving on the editorial board of the Journal of Biomedical Informatics and Chairing the 2022 American Medical Informatics Association (AMIA) Informatics Summit. Further she has co-developed the Coursera Clinical Data Science Specialization - a series of six MOOCs designed to teach clinical research informatics skills.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: {ReviewR}: A Shiny App for Reviewing Clinical Records  
## abstract: Reviewing clinical records is essential for data extraction and validating algorithms. Traditionally, researchers review the electronic health record (e.g., Epic, Cerner) and record abstractions in spreadsheets (e.g., Excel). However, this has workflow challenges (multiple windows, copy/pasting wrong patient identifiers, etc.), and increasingly organizations are limiting EHR access for non-clinical researchers. To solve these challenges, we created ReviewR, a Shiny application that combines the display of patient records from clinical data warehouses with an abstraction interface to improve workflow and review accuracy. ReviewR is architected to be flexible across clinical data models (OMOP, MIMIC-II) and databases (Google BigQuery, Postgres). We include functions to help users extend ReviewR to support custom data models and databases as needed. Chart view functionality is supported by the DT package, allowing each table to be filtered by multiple columns and text searching, filtering, and highlighting. The chart abstraction interface allows for multiple types of data capture (e.g., radio buttons, check lists, free text entry, etc.) uploaded securely to the research data management system REDCap. ReviewR is available on CRAN.  
##   
##   
## Your Score (1-9):   
##   
## name: Andreas Soteriades   
## org: Nottinghamshire Healthcare NHS Foundation Trust  
## title: Data Scientist   
## time zone: (GMT) Greenwich Mean Time: Dublin, Edinburgh, Lisbon, London  
## bio: I am an Applied Mathematician with Data Science experience in both academic and non-research areas, from dairy farm efficiency modelling to Text Mining (plus a few other in between!). I am currently crunching NHS patient feedback data (text classification and sentiment analysis) and translating findings into an illuminating dashboard for managers and non-technical staff.  
## gender: NA  
## poc: NA  
## submission\_type: regular  
## abstract\_title: Scaling up: Deployable Shiny and Text Mining to Guide Decision-Making in the Healthcare Sector of an Entire Country  
## abstract: Undoubtedly, R is a powerful tool for data manipulation and visualization, statistical analysis, and machine learning (ML). Shiny is excellent for producing dashboards, and so are “tidymodels” and “mlr3” for building ML pipelines. However, Shiny developers often come across situations where the code for a complicated Shiny app becomes massive, hard to read, understand and debug, and it is so tailored to the dataset that the app cannot work as a framework for other datasets. Similarly, ML practitioners often face the fact that R simply cannot handle efficiently large ML pipelines and datasets. In this talk, we will discuss how we built and deployed a scalable and robust text classification pipeline and dashboard using the R packages “Golem” and “reticulate”. Golem helps build and ship Shiny apps as R packages that are modular, agnostic to deployment and submission-ready (performs tests automatically). Package “reticulate” allows running Python through R, which gave us access to the state-of-the-art Scikit-learn Python library for ML. We will show how we successfully used both packages to develop a tool for England’s National Health Service.  
##   
##   
## Your Score (1-9):   
##   
## name: Judith Thomas Lewis   
## org: Vanderbilt University Medical Center  
## title: Application Developer/Adjoint Assistant Professor   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Judy Lewis is a biomedical informatics application developer at Vanderbilt University Medical Center. As the lead developer of the Harmonist Data Toolkit, she is focused on creating generalizable software for observational medical data and leveraging REDCap in reproducible workflows. After transitioning to R development after 20+ years of teaching biomedical engineering undergraduates, Judy was happy to find the collaborative, welcoming R community, share her work at local RLadies meetings, and take part in RLadies gatherings at rstudio::conf. Judy earned her doctorate in biomedical engineering at Vanderbilt University.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: R/Shiny and REDCap: Streamlining Data Workflow in a Global Research Consortium  
## abstract: We will describe the design and impact of a web-based R/Shiny/RMarkdown application that provides a data harmonization workflow for large observational research consortia. This cloud-based application has an advanced Shiny UI and is integrated with REDCap via the R httr package. Users can upload datasets to check data quality, generate reports, visualize data, and securely transfer files. Details about the expected format of the data (tables, variables, formats, valid codes for coded variables defined in the data model) are stored in REDCap and accessed via the R/REDCap API. The application uses these details to auto-generate data quality checks and generate customizable PDF reports for download. User authentication, file transfer information, and audit trails are also exchanged with REDCap.  
## One year after the tool was launched in an international HIV research consortium, users reported that the data quality feedback, reproducible reports, and easy-to-use workflow of the Harmonist Data Toolkit helped them improve data quality and decreased the amount of time required for data preparation and reporting. The Toolkit is designed to be adapted for use by other research networks.  
##   
##   
## Your Score (1-9):   
##   
## name: Jonathan Trattner   
## org: Wake Forest School of Medicine  
## title: Master's Student   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Hi! I’m Jonathan Trattner. I recently completed a B.S. in Computational Neuroscience from Wake Forest University in Winston-Salemn, North Carolina. I’m now a Master’s student in the Kishida Lab at Wake Forest School of Medicine. My primary research interests involve human decision-making, specifically neuroeconomics. I’m an experienced #rstats programmer who dabbles in web development, Bayesian Statistics, Reinforcement Learning, and #dataviz.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Automating Human Behavioral Data Collection and Analysis with {shiny}  
## abstract: Collecting human behavioral data follows a common pattern where data is collected first and subsequently analyzed. Recent clinical trials have used Bayesian methods to continually analyze data and improve study designs, saving time, resources, and lives (Freeman et al., 2020).   
##   
## In line with this practice, I introduce a suite of packages, {shinyresearch}, to help human behavioral researchers more efficiently analyze their data and, if needed, adapt their study. This suite is built around {shiny}, which integrates interactive web applications with R, and allows for continual and automated data analysis. This is facilitated with three packages used in an ongoing study approved by the institutional review board (IRB) at Wake Forest School of Medicine.   
##   
## {shinyland} helps researchers easily create task ‘landing pages’ to describe study goals and collect informed consent; {shinyrandomize} provides helper functions for generating unique ids and splitting participants into cohorts; {shinysurveys} allows for easily collecting questionnaire data with helper functions for aggregating responses in a tidy format for easy analysis.  
##   
##   
## Your Score (1-9):   
##   
## name: Erika Helgeson   
## org: University of Minnesota  
## title: Assistant Professor   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Erika Helgeson is an Assistant Professor in the Division of Biostatistics at the University of Minnesota. Her research interests include statistical machine learning and nonparametric methods for complex, high-dimensional data. She collaborates extensively with investigators in transplantation and COPD and has additional research interests in chronic pain conditions.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: NA  
## abstract: Individuals considering kidney donation would benefit from understanding their personalized long-term health risks prior to undergoing this life changing medical procedure. Peer reviewed articles may provide cutting edge research about such risks but are often not accessible to the donor candidate nor can the candidate easily determine their personalized risk from these articles. In our presentation, we illustrate how R Shiny can be used to fill this gap: translating scholarly findings into an understandable, user-friendly application.   
##   
## We present our work on visualizing long-term risk models for kidney donors. While calculators for donor risk of hypertension, diabetes, proteinuria, and reduced renal function have been published, these calculators are either only presented as model summaries or require familiarity of Excel to generate personalized risk estimates. To increase accessibility for potential donors and their care team, we adapted these risk models into a user-friendly web-based calculator using Shiny. ,This risk assessment tool provides a comprehensive long-term projection of living kidney donor health in an easy to use and understandable format.  
##   
##   
## Your Score (1-9):   
##   
## name: Michael Kane   
## org: Yale University  
## title: Assistant   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Michael Kane is an Assistant Professor of Biostatistics at Yale University. His research focuses on areas of machine learning, computing, and reproducibility in application areas including enrollment heterogeneity in clinical trials and population-scale human mobility.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Using Shiny to Visualize a Kidney Donor’s Personalized Long-Term Risk  
## abstract: Clinical trial data, recorded as ADaM (or ADaM-like) data, attempts to capture all measurements made during a clinical trial that may be used to assess patient prognosis and the predictive relationships between a therapy and a set of outcomes. Biostatisticians’ whose job is to analyze these types of data collaborate with clinicians to understand the data through tables, visualizations, and preliminary analyses. Even though these data are structured they may still be complex and require iteration between biostatisticians and the clinicians with whom they are collaborating. In this talk, we will show how several R packages can be used in conjunction to create more comprehensive exploration to reduce this iteration. The approach has users assign roles to variables and then automates the process of generating summaries of the bivariate relationships between variables in specified roles. These summaries can then be displayed in a navigable, interactive html environment that can be disseminated to all members of a collaboration for review and discussion. We will provide several examples of these reports and show how they can be extended to include analyses thereby encapsulating all stages of the data analysis.  
##   
##   
## Your Score (1-9):   
##   
## name: Eric Brown   
## org: University of Toronto  
## title: Assistant Professor   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Dr. Eric Brown is a psychiatrist and researcher at the Centre for Addiction and Mental Health in Toronto, Canada and an Assitant Professor in the Department of Psychiatry, University of Toronto. His research is focused on geriatric psychiatry including risk factors for dementia and cognitive impairment, neuroimaging, environmental contributors to cognitive decline, and more recently, COVID-19 and its impacts on mental health.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: A Shiny-based online calculator to estimate the risk of undetected COVID-19 in a congregate setting following an exposure  
## abstract: High rates of community transmission of COVID-19 lead to possible exposures among patients or residents of congregate settings. The resultant necessary increase in infection prevention and control (IPAC) measures have potential harms such as loneliness and deconditioning. An evidence-based estimation of the risk of the presence of undetected COVID-19 may help to optimize the timing of IPAC measures. We report an innovative Shiny-based online calculator and R package to estimate the probability of an undetected COVID-19 case within a given setting following a potential exposure when there are no symptomatic patients and polymerase chain reaction (PCR) testing is available. The calculator incorporates estimates of incubation period, PCR sensitivity by day of testing, and the estimated fraction of true asymptomatic cases. Our calculator can facilitate evidence-based decisions in a clinical setting or enhance the development of guidelines. The underlying code is publicly available for transparency and accessibility.  
##   
##   
## Your Score (1-9):   
##   
## name: Peter Higgins   
## org: University of Michigan  
## title: Professor   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I teach R to fellows and junior faculty interested in doing reproducible medical research. I also do research in inflammatory bowel diseases, machine learning, biomarkers, drug development, epidemiology, photoacoustics, adaptive clinical trials as part of my clinical practice in gastroenterology, with the support of 3 NIH R01 grants.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Using the {medicaldata} package for teaching #rstats  
## abstract: It can be difficult to find good datasets for teaching medical research. Many are proprietary or hidden behind privacy concerns. But learners really appreciate working with real medical data, particularly from clinical trials. I have gathered together several datasets, reconstructed others, and received some lovely donations of deidentified data sets.  
##   
## These have been packaged into the {medicaldata} package, which is available on Github, and if I am especially stubborn and persistent, possibly soon on CRAN. This package includes several well-documented datasets, from as early as 1757 (James Lind and scurvy) to 2020 (COVID testing). Several of the clinical trials were ones I was involved with, and others were reconstructed from tables (Streptomycin for Pulmonary TB).   
## This package can be a helpful teaching resource, as the datasets are carefully documented, with explanations and codebooks, both in the help files and online with links to pdf documents on each dataset in the Github repo at https://github.com/higgi13425/medicaldata.   
## This package functions as a companion to the e-book Reproducible Medical Research with R, but can be used independently. https://bookdown.org/pdr\_higgins/rmrwr/.  
##   
##   
## Your Score (1-9):   
##   
## name: Sujata Patil   
## org: Cleveland Clinic Foundation  
## title: Biostatistician   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I am Section Head of the Cancer Biostatistics group at Cleveland Clinic, a position I started in January 2021. Before that, I spent 15 years at Memorial Sloan Kettering Cancer Center as a collaborative biostatistician. My interests are in study design, survival analysis, and missing data analysis.  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Using free interactive applications to communicate statistical concepts to laboratory researchers  
## abstract: Statistics literacy is crucial for pre-clinical laboratory researchers engaged in anticancer drug development. We developed a short course in statistics for biomedical researchers using didactic lectures and data examples from published research. Yet, foundational concepts often remain abstract to most researchers. Therefore, we developed computer-based apps that allow biomedical researchers to explore and visualize fundamental ideas at their own pace, thus empowering them to understand and use statistics in their research.  
##   
## The interactive computer apps covered various concepts, including experimental design, p-values, and multiple comparisons. Our free R Shiny apps facilitate inquiry-led activities to increase motivation and promote statistical thinking and can used by interested colleagues from anywhere in the world, thus democratizing statistics education.  
##   
## We evaluated the computer apps to assess how well participants were able to comprehend statistical concepts from the beginning to the end of the course and if the apps were useful. We will present our apps as well as results from these evaluations.  
##   
##   
## Your Score (1-9):   
##   
## name: Jie Cao   
## org: University of Michigan  
## title: PhD Student   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: NA  
## gender: female  
## poc: yes  
## submission\_type: lightning  
## abstract\_title: clinspacy: An R Package for Clinical Natural Language Processing  
## abstract: Clinical text contains rich information that is potentially powerful input for machine learning in medicine. However, natural language processing (NLP) tools used to process clinical text need to be tailored to the clinical context. Building on Python’s spaCy framework, scispaCy’s clinical language model and Unified Medical Language System (UMLS) concept mapping, and medspaCy’s context and sectionizer tools, clinspacy provides a single-function capability to process massive volumes of clinical text directly to a data frame or a CSV file.  
##   
## Although clinspacy depends upon Python, users do not need to install it manually. The package automatically detects and installs miniconda and all dependencies. Beyond providing an interface to the aforementioned Python packages, clinspacy also enables the creation of model-ready datasets through its bind\_\*() functions that can be used to create bag-of-entities, bag-of-concepts, and concept-embedding datasets leveraging either scispaCy or cui2vec embeddings. clinspacy is available on GitHub and CRAN.  
##   
##   
## Your Score (1-9):   
##   
## name: Taylor Arnold   
## org: University of Richmond  
## title: Associate Professor of Statistics and Linguistics   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Taylor Arnold is an Associate Professor of Statistics and Linguistics at the University of Richmond. A recipient of grants from the NEH, ACLS and the Mellon Foundation, Arnold's research focuses on computational statistics, text analysis and image processing. His first book Humanities Data in R (Springer, 2015) explores four core analytical areas applicable to data analysis in the humanities: networks, text, geospatial data, and images. His second book, A Computational Approach to Statistical Learning (CRC Press 2018), explores connections between modern machine learning techniques with theories of statistical estimation. Numerous journal articles extrapolate on these ideas in the context of particular applications. Arnold has also released dozens of open-source libraries in R, Python, Javascript and C. Prior appointments include positions at AT&T Labs Research, IBM, Universtié Paris Diderot, and the Institut d'Études Avancées at ENS-Lyon.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: ctrialsgov: Access, Visualization, and Discovery of the ClinicalTrials.gov Database  
## abstract: In this talk I will present our R package ctrialsgov, which provides functions for accessing and visualizing data from ClinicalTrials.gov’s open access database. The package focuses on providing an easy-to-use interface to clinical trial data along with useful text analysis functions for working with the rich free text response fields present in the dataset. The lightening talk will focus on introduction the package through a case study that investigates how the package could be used to automatically find and sort innovative clinical trial designs for cancer research. The R package is released under an MIT license and currently installable from GitHub.  
##   
##   
## Your Score (1-9):   
##   
## name: Irene van den Broek   
## org: JeBentWatJeMeet  
## title: Freelance data visualizer   
## time zone: (GMT+01:00) Amsterdam, Berlin, Bern, Rome, Stockholm, Vienna  
## bio: Irene van den Broek (PhD) has a background as clinical research scientist in the field of proteomics and laboratory medicine. In the past two years, she has shifted gears from molecular to digital biomarkers and specialized in the analysis and visualization of health data. With her freelance business - You are what you measure – she combines her passions for self-quantification and data visualization with her interest in personalized health and biomarker research. Through the creation of (interactive) data visualizations and (R Shiny) web applications she supports researchers and clinicians to make health data more actionable, accessible and interpretable.  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: You R What You Measure: Digital biomarkers for insights in personalized health  
## abstract: Wearable devices enable continuous, longitudinal collection of physiological or behavioral data. Initially geared towards consumer health enthusiasts, research- and clinical grade wearable devices now find their way into medical research and clinical trials. This talk will highlight the potential of digital biomarkers to advance personalized medicine as well as the central role for R in the connection, integration, analysis, and/or visualization of health data from digital devices.  
## First, two digital biomarkers with promising clinical utility will be introduced: continuous glucose monitoring (CGM) and heart rate variability (HRV). The discussion will also include personal use experiences and reference to R packages specifically developed for CGM and HRV data analysis. The second part of the talk will showcase examples from a two-year long personal project (you R what you measure) to quantify and visualize health and habits. Use cases include the connection and visualization of data from popular wearable devices and smartphone apps for sleep, activity and productivity tracking, all using the R programming language.  
##   
##   
## Your Score (1-9):   
##   
## name: Travis Gerke   
## org: The Prostate Cancer Clinical Trials Consortium  
## title: Director of Data Science   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Travis Gerke, ScD, is Director of Data Science at The Prostate Cancer Clinical Trials Consortium, a multicenter clinical research organization. His academic background intersects biostatistics, epidemiology, and bioinformatics. As a data scientist, Dr. Gerke is a long-time R enthusiast who enjoys leading teams towards adoption of modern data-related workflows that accelerate and fortify the oncology research pipeline.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: CONSORT diagrams in R with {ggconsort}  
## abstract: CONSORT diagrams are the industry standard graphic for representing participant flow in a clinical trial. Several teams have considered R packages for CONSORT diagram construction, but have fallen short of end-to-end automation. Most notably, Dr. Peter Higgins proposed an excellent CONSORT design overview in his R/Medicine 2020 talk, in which he described the legacy diagram construction as "an artisanal product built by counting categories, then copy and pasting results into templates." Here, I will present a fresh look at the {ggconsort} package, which embraces aspects of CONSORT construction that are inherently artisanal, but eliminates error-prone manual entry. Specifically, I segment CONSORT creation into two stages: (1) CONSORT annotation capture at the time of data wrangling, and (2) diagram layout. I will demonstrate how stage (1) can be built into a single tidyverse chain. Stage (2) maintains some artisanal aspects, but {ggconsort} provides new, helpful geoms to streamline the process.  
##   
##   
## Your Score (1-9):   
##   
## name: Jared Huling   
## org: University of Minnesota  
## title: Assistant Professor of Biostatistics   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Jared Huling is an Assistant Professor in the University of Minnesota School of Public Health. His research interests focus on the development of precision medicine, causal inference, and statistical learning methodology for the analysis of complex observational studies. He is particularly interested in addressing various forms of population heterogeneity with the aim of improving patient health outcomes. His work in this area has involved applications in health system risk modeling and in personalizing health system intervention enrollment decisions and involves delivering open source statistical software implementations of his work.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Subgroup identification and precision medicine with the 'personalized' R package  
## abstract: Heterogeneity of treatment effect is common in medical studies. Improving medical decision making by matching patients with the most effective treatments has the potential to improve treatment efficacy and adherence. Exploratory subgroup analysis allows the identification of clinically relevant subgroups from pre-specified variables for such purpose. In this talk, we will introduce a general framework that encompasses many recent statistical methods for identifying subgroups of patients who may benefit from different available treatments and an R software package, personalized, which implements this general framework. The methods in this framework estimate optimal individualized treatment rules (ITRs) that match patients with treatments by focusing on interactions between the treatments and covariates. We will cover the analysis of data from both randomized clinical trials and observational studies and a wide range of types of outcomes including continuous, binary, count, and times-to-event. This talk will illustrate the usage of this package, from estimation of ITRs for various outcomes to validating and visualizing whether the estimated ITRs result in improved outcomes when applied to new patients.  
##   
##   
## Your Score (1-9):   
##   
## name: Emily Zabor   
## org: Cleveland Clinic  
## title: Assistant Staff Biostatistician   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: Emily works as a cancer biostatistician, with a focus on the development of statistical methods for the design of early phase oncology clinical trials and the analysis of observational data.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: Designing early phase clinical trials with the {ppseq} package  
## abstract: This talk will introduce the {ppseq} R package for designing early phase clinical trials using sequential predictive probability monitoring. Clinical trials in oncology increasingly include dose-expansion cohorts to further characterize safety, get preliminary efficacy data, or compare across doses or disease subtypes. But expansion cohorts are often added to trials on the fly, with little planning of the statistical design or consideration for the ethical concerns associated with treating patients at sub-optimal or inefficacious doses. The {ppseq} package provides functionality to design trials using Bayesian sequential predictive probability monitoring. The package includes interactive visualizations that allow users to easily compare designs based on different decision thresholds according to their operating characteristics, and introduces optimization criteria to assist in selecting the ideal design. The functionality of the {ppseq} package will be demonstrated with a re-design of the phase 1 dose-expansion cohort studying the anti-PD-L1 treatment atezolizumab in metastatic urothelial carcinoma.  
##   
##   
## Your Score (1-9):   
##   
## name: Brandon LeBeau   
## org: University of Iowa  
## title: Associate Professor   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Brandon LeBeau is an Associate Professor of Educational Measurement and Statistics at the University of Iowa. His research interests include quantitative program evaluation, longitudinal data methodology, and research software development. He has created a handful of R packages used for research purposes, including pdfsearch and simglm.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Tidy Simulation and Power Analyses  
## abstract: Power analyses are commonly needed to secure external funding from Federal, State, or local agencies. Many current software implementations for power analyses make assumptions in order for the calculations to be mathematically possible. Unfortunately, the data collected commonly do not meet the assumptions typically made within software used for power analyses. A more flexible and possibly more accurate power analysis framework can be achieved through Monte Carlo simulation. This presentation aims to introduce the benefits and flexibility achieved when conducting a power analysis under a Monte Carlo framework. The presentation will focus specifically on power analyses for general(-ized) linear (mixed) models using an R package called simglm (https://github.com/lebebr01/simglm). This package allows for flexible specification of data simulation conditions that may better approximate real data conditions. In addition, the package allows for the varying of data generating conditions, model fitting conditions, and others to explore impact of specific data characteristics on the empirical power.  
##   
##   
## Your Score (1-9):   
##   
## name: Emil Hvitfeldt   
## org: Teladoc Health  
## title: Data Analyst   
## time zone: (GMT-08:00) Pacific Time (US and Canada); Tijuana  
## bio: Emil Hvitfeldt is a clinical data analyst working in healthcare, and an adjunct professor at American University teaching statistical machine learning with tidymodels. He is also an open source R developer.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Creating and styling pptx slides with rmarkdown  
## abstract: {rmarkdown} is a wonderful package that allows us to create many types of outputs. From simple pages to slides, dashboards, and websites. Even though {xaringan} slides have gained a lot of popularity, there is sometimes the need to create slides in Microsoft PowerPoint format. This talk will show you how to create simple .pptx files with {rmarkdown} and how they can be styled to conform with your organization's style. This talk is a formal introduction to the {pptxtemplates} package which contains examples of styled .pptx templates.  
##   
##   
## Your Score (1-9):   
##   
## name: Kristen Panthagani, PhD   
## org: Baylor College of Medicine  
## title: MD-PhD trainee   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I am a senior MD-PhD student at Baylor College of Medicine who runs the science communication blog youcanknowthings.com. I learned R for my thesis work studying the human microbiome, and over the last year have used my R skills to create animated visualizations of COVID data as a science communication tool. I am in my final year of medical school and will be applying to residency in Emergency Medicine this fall.  
## gender: female  
## poc: no  
## submission\_type: lightning  
## abstract\_title: Animated data visualizations with gganimate as a science communication tool during the pandemic  
## abstract: Unlike any time in recent memory, the challenges, uncertainties, and misinformation surrounding the COVID pandemic have caused the general public to be highly interested in understanding data for themselves. To help address this need, in March 2020 I began creating data visualizations to help the general public understand the growth of the pandemic. Using ggplot2 and gganimate, I create data visualizations that have since been viewed hundreds of thousands of times across multiple social media platforms. Several features make data gifs particularly effective as science communication tools, including rescaling of axes over time to illustrate exponential growth, creating animated data labels with ggrepel to make graphs very easy to understand, and using beautiful animation as a way to capture attention and increase engagement. In my talk I will briefly highlight how to use these features in R and why they are effective science communication tools.  
##   
##   
## Your Score (1-9):   
##   
## name: Andy South   
## org: afrimapr - Liverpool School of Tropical Medicine  
## title: Data Scientist   
## time zone: (GMT) Greenwich Mean Time: Dublin, Edinburgh, Lisbon, London  
## bio: R developer leading the afrimapr project for past 18 months. Committed to R, code sharing, open-source & open-data as more efficient means of addressing health issues in both low & high income countries. Lover of maps & datavis.  
## gender: male  
## poc: no  
## submission\_type: lightning  
## abstract\_title: mapping African health data with afrimapr packages, training & community  
## abstract: The afrimapr project is supporting analysts in Africa to make maps from health data by 1. developing R packages & components 2. providing open-access training materials 3.  
## initiating a community of users and developers. The R packages can be used to create data visualisations, reproducible reports, and web applications. This includes the ability to make maps from data that are referenced by coordinates or place names. Local data-analysts can be supported to develop or modify their own solutions to  
## local issues. Data science & R communities are rapidly growing in Africa. R components can be linked to each other and other data systems. Our work eases access to available open data on health facility locations, admin boundaries, sub-national covid case data and population estimates. We conducted the most recent and reproducible review of open access health facility location data for African countries. Demonstration web applications and training materials are available from the afrimapr.org website. In this talk I will demonstrate the range of our packages, training materials and initial community building activities. We welcome contributions.  
##   
##   
## Your Score (1-9):   
##   
## name: Steven Schwager   
## org: Cornell University, Dept of Statistics & Data Science  
## title: Professor Emeritus   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: After earning my PhD in Statistics from Yale, I joined the Biometry and Statistics faculty at Cornell. My research has addressed problems in biostatistics, ecological statistics, outlier detection, epidemiology, experiment design, modeling, multivariate analysis, sampling design, statistical computing, statistical data analysis, and other areas. I have worked on developing and applying statistical methods for the biological, physical, and social sciences, engineering, insurance, and clinical trials.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Graphical Displays in R for Clinical Trials  
## abstract: Graphical display of clinical trial data can be much more informative than numerical tables and lists. Visualization of study data is an essential tool for evaluating the trial’s progress and evaluating and understanding its outcome. However, creating effective graphics involves art as well as science, requiring careful thought and execution, as is well known. We illustrate a few pitfalls that must be avoided and guidelines for well-designed, powerful graphics, citing the seminal work of Tufte, with focus on clinical trial data. The versatility of R makes it well-suited to creating graphics for exploring the data; monitoring and assessing safety and efficacy; displaying statistical variability; detecting outliers and other anomalies; summarizing the data clearly and correctly; and communicating results to a wide audience, including both regulators and non-statisticians. We present an overview of graphics for clinical trials, including major types of displays, with illustrative examples. We show how R can be used to create innovative graphics by combining approaches and components in novel ways to achieve specialized goals.  
##   
##   
## Your Score (1-9):   
##   
## name: Thomas Nowlan   
## org: Adelphi Research Global  
## title: Senior Project Director / Advanced Methods Associate   
## time zone: (GMT-05:00) Eastern Time (US and Canada)  
## bio: I have worked in a variety of positions within the healthcare system, and I use my wide range of experiences to help inform my decisions as a quantitative researcher. I am passionate about leveraging the best methods available to explore the pharma/healthcare industry. More recently, I have utilized R studio to bring to life meaningful analyses through automated processes and highly engaging graphics.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Visualizing Treatment Pathways in R  
## abstract: Understanding the complexities of how health care providers (HCPs) prescribe treatments over time is an area of great interest to health care market researchers. Fundamental to this understanding is the examination of treatment pathways, the order in which HCPs typically prescribe indicated therapies to appropriate patients. Historically, this analysis involves a fair amount of data cleaning and is challenging to visualize into one clean output.   
##   
## This talk will examine how R can be leveraged to create intuitive charts to display treatment pathways. By utilizing the R package ggalluvial, it is now possible to create highly engaging outputs to examine prescribing sequences. In the new format, it is easier to understand the multitude of ways HCPs may move through treatments. Furthermore, the code required to create the charts is concise and the outputs are highly adaptive and reproducible. These outputs can help identify the positioning of treatments, what mechanisms of action are favored, and help identify different HCP segments based on prescribing behavior.  
##   
##   
## Your Score (1-9):   
##   
## name: Andrew Bland   
## org: Grand River  
## title: Nephrologist   
## time zone: (GMT-06:00) Central Time (US and Canada)  
## bio: Dr. Bland has training and experience in the research, development and deployment of Digital Health solutions across several healthcare enterprises including using predictive analytics, NLP, and data visualization to improve value-based care and reduction of healthcare related adverse events.  
##   
## Dr. Bland is triple boarded in Internal Medicine, Pediatrics, and Nephrology. He has a Masters of Business Administration from the Univeristy of Massachusetts and a Masters of Data Science and Predictive Analytics from Northwestern. In 2020 he was recognized by Becker's Health Care as one of the 100 Chief Medical Officers to Know.  
## gender: male  
## poc: no  
## submission\_type: regular  
## abstract\_title: Improving Quality and Safety of Dialysis Patients with Control Chart Visualization of Time Series Data  
## abstract: Patients with End Stage Renal Disease (ESRD) who require dialysis to live are 1.3% or $49.2B of the Medicare population, yet consume 7.2% of the annual $507.9B Medicare spend (Annual Data Report, n.d.)Despite this exorbitant consumption of resources, the value proposition, , of dialysis care is significantly lower than other chronic health conditions (Polkinghorne, 2019).  
##   
## ESRD data sources include; monthly comprehensive labs, twelve hours of supervised care in a dialysis center including every thirty minute vital signs, four time a month physician visits, monthly required meetings with dietitians and social workers, and required quarterly quality of life surveys. Transformation of this large data set into actionable information that can be given in near realtime to the care team and patients must be a priority if the value of care is to be improved (Lacson et al., 2009).   
##   
## Transformation of raw data to tidy-data based visualizations of time series data and control charts may help patients and their care team understand the patient journey through time rather than a series of snapshots of individual months data presented without the context of time.  
##   
##   
## Your Score (1-9):   
##   
## name: Julia Silge   
## org: RStudio PBC  
## title: software engineer   
## time zone: (GMT-07:00) Mountain Time (US and Canada)  
## bio: Julia Silge is a data scientist and software engineer at RStudio PBC where she works on open source modeling tools. She is an author, an international keynote speaker, and a real-world practitioner focusing on data analysis and machine learning practice. Julia loves text analysis, making beautiful charts, and communicating about technical topics with diverse audiences.  
## gender: female  
## poc: no  
## submission\_type: regular  
## abstract\_title: Data visualization for machine learning practitioners  
## abstract: Visual representations of data inform how machine learning practitioners think, understand, and decide. Before charts are ever used for outward communication about a ML system, they are used by the system designers and operators themselves as a tool to make better modeling choices. Practitioners use visualization, from very familiar statistical graphics to creative and less standard plots, at the points of most important human decisions when other ways to validate those decisions can be difficult. Visualization approaches are used to understand both the data that serves as input for machine learning and the models that practitioners create. In this talk, learn about the process of building a ML model in the real world, how and when practitioners use visualization to make more effective choices, and considerations for ML visualization tooling.  
##   
##   
## Your Score (1-9):

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.