



# **DairyCoPilot - Automated Data Compilation and Analysis Tools for DairyComp Data Assets**

Srikanth Aravamuthan, Dörte Döpfer, and Emil Walleser  
Department of Medical Sciences, School of Veterinary Medicine,  
University of Wisconsin, Madison 53706  
[walleser@wisc.edu](mailto:walleser@wisc.edu)

# Dairy Farm Management

- Data-driven decisions
- Successful farms record and analyze data
  - Financial
  - Input
  - Production/health
- Animal health
  - Need data and tools



# DairyComp

- Dairy management software
- Interconnected apps
- Most popular software



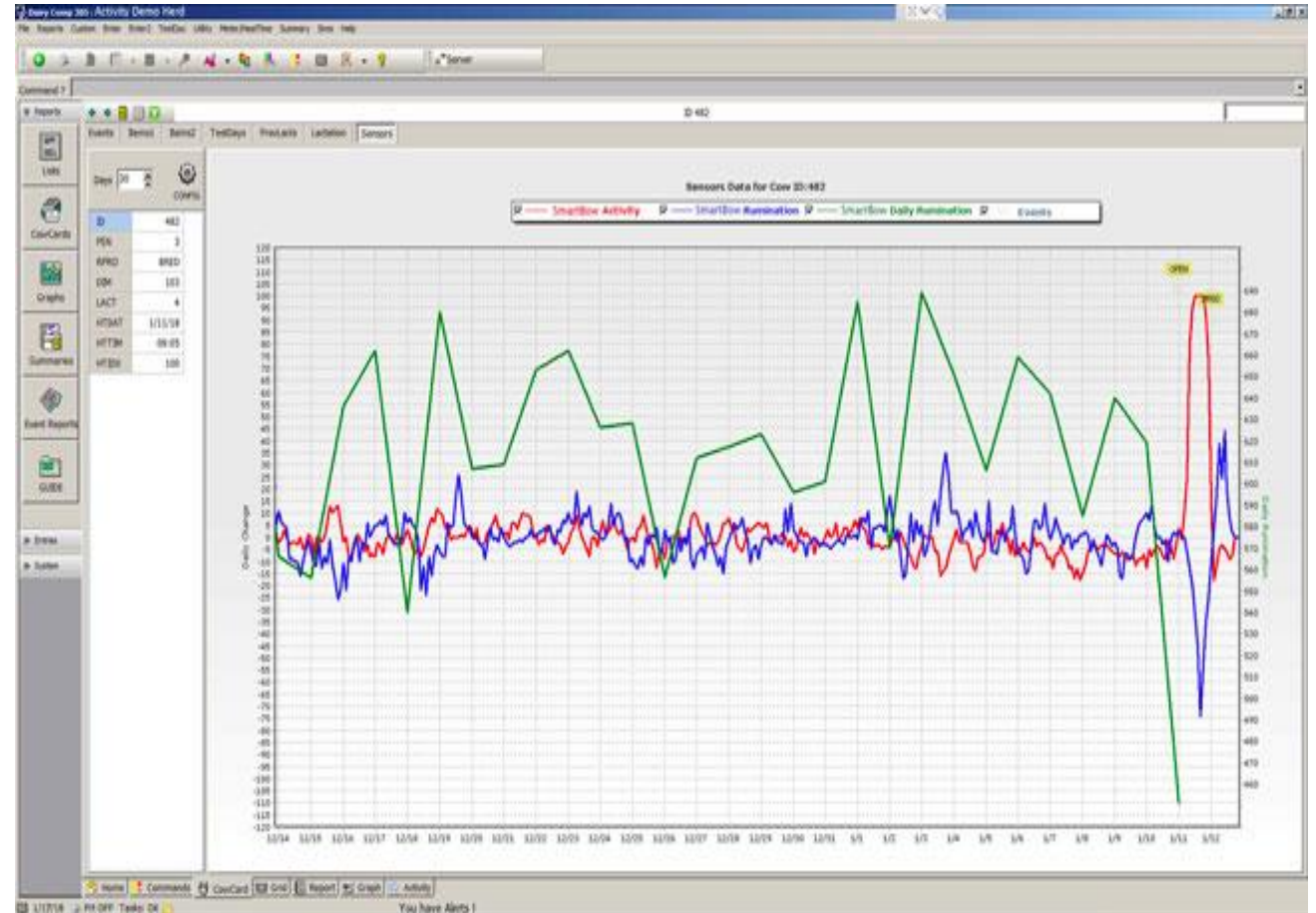
# Strengths

- Large user base
- Multiple integrations
- User customizability



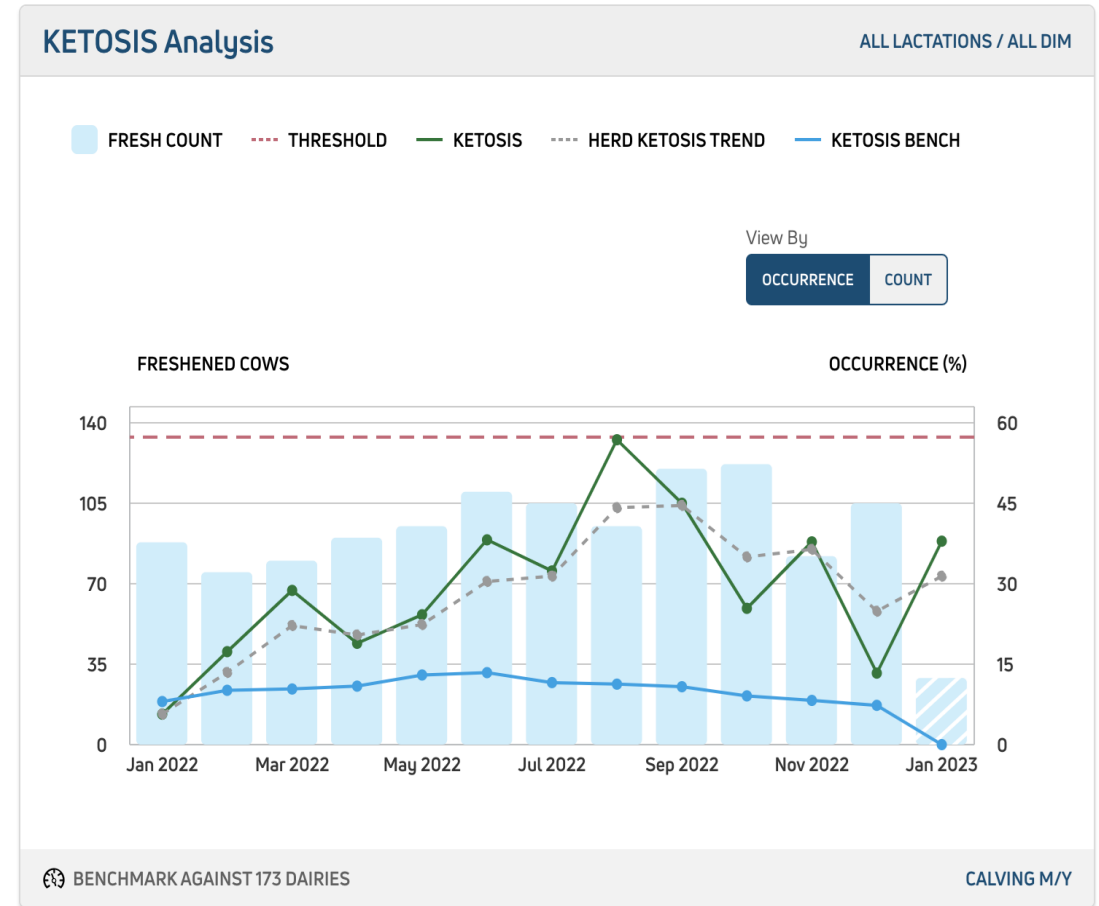
# Weaknesses

- User customizability
- Aged desktop software
- Poor analysis options
- Proprietary software



# Motivation

- Health data is locked in DC
- Limited integrated analysis
  - Trend lines
  - Benchmarking
- Extensive preprocessing





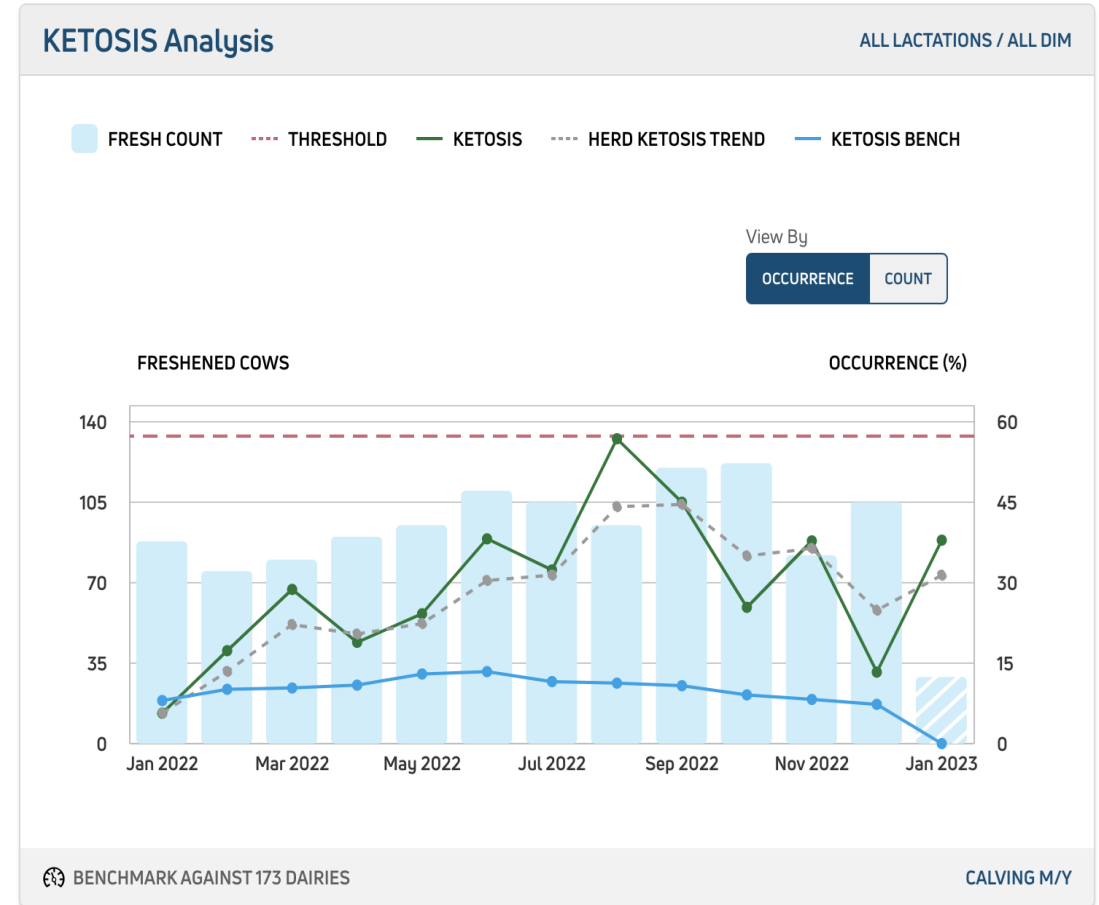
# Automate DairyComp

## Aim

- Automate DairyComp write out
- Apply tools to support data-driven decisions
- Customize reports and graphs for a better management experience

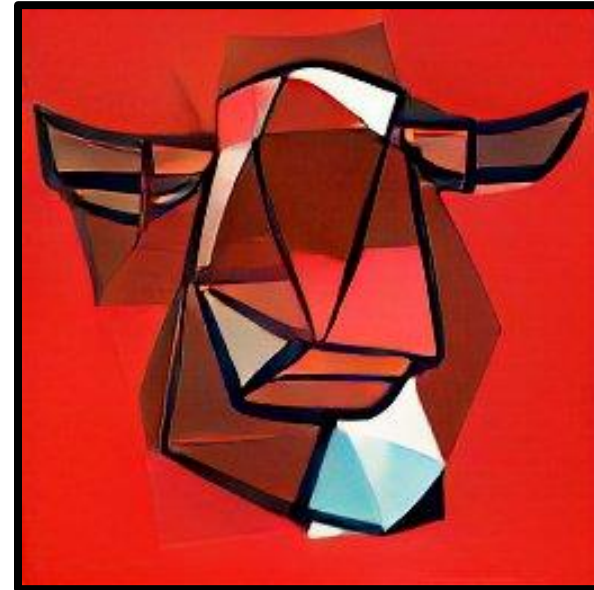
## Audience

- DairyComp users
- Dairy Skills I instructors & students



# DairyCoPilot

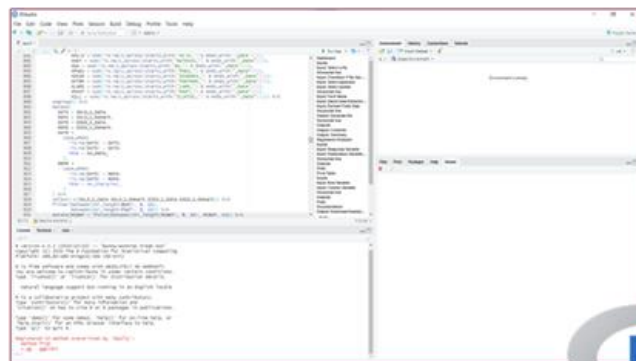
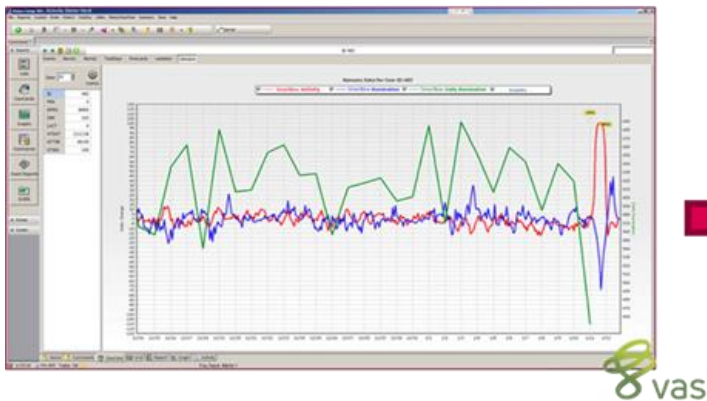
- Simplify data wrangling in DairyComp
  - Time spent creating events < 20 mins
  - Repeatability
- Reduce time for data cleaning
  - Download → Insight < 10 mins
- Simplify analysis
  - GUI
  - Interactive
  - Reportability
- Enhanced analysis
  - Multiple odds ratios → Multiple linear regression





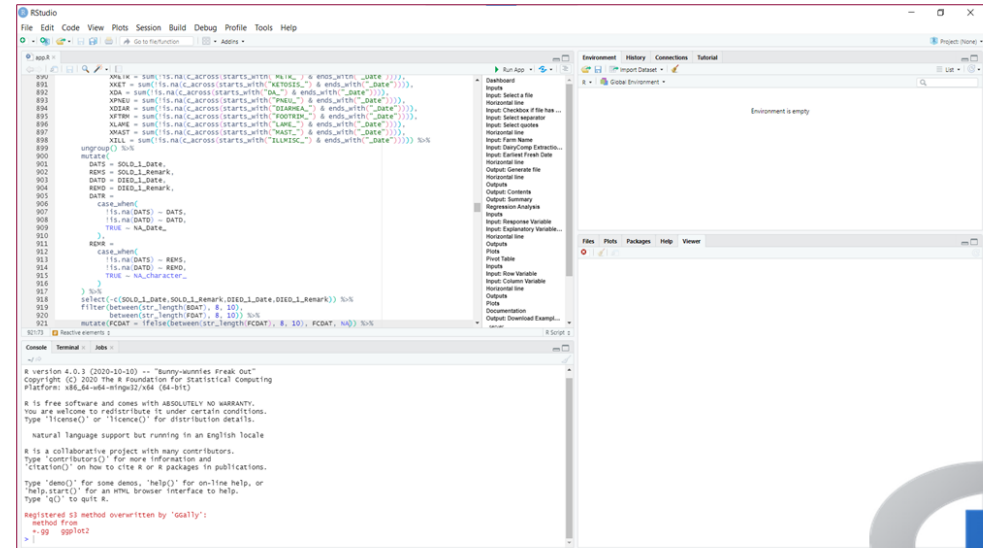
# Methods

1. DairyComp Data Extraction
2. DairyCoPilot Data Cleaning
3. Rstudio Connect Deployment



# DairyCoPilot Data Cleaning

- Read CSV file
  - Rename events
  - Filter out missing events
  - Create missing columns
- Delete duplicate event records
- Create predefined variables
  - Number of occurrences of an event
  - Date and Remark of first event
  - Cutoff days post calving
- Convert to appropriate variable type



The screenshot shows the RStudio interface with a script editor containing R code for data cleaning. The code includes comments in Chinese and R functions for reading a CSV file, renaming columns, filtering out missing values, and creating new variables. The console at the bottom shows the R version and system information.

```
# 读取CSV文件
# 1. 读取CSV文件
# 2. 重命名事件
# 3. 过滤掉缺失事件
# 4. 创建缺失列
# 5. 删除重复事件记录
# 6. 创建预定义变量
# 7. 转换为适当的变量类型

# 读取CSV文件
data <- read.csv("data.csv")

# 重命名事件
colnames(data) <- c("date", "event", "remark")

# 过滤掉缺失事件
data <- data[!is.na(event), ]

# 创建缺失列
data$case_name <- NA
data$case_date <- NA
data$case_remark <- NA

# 删除重复事件记录
data <- data[!duplicated(data), ]

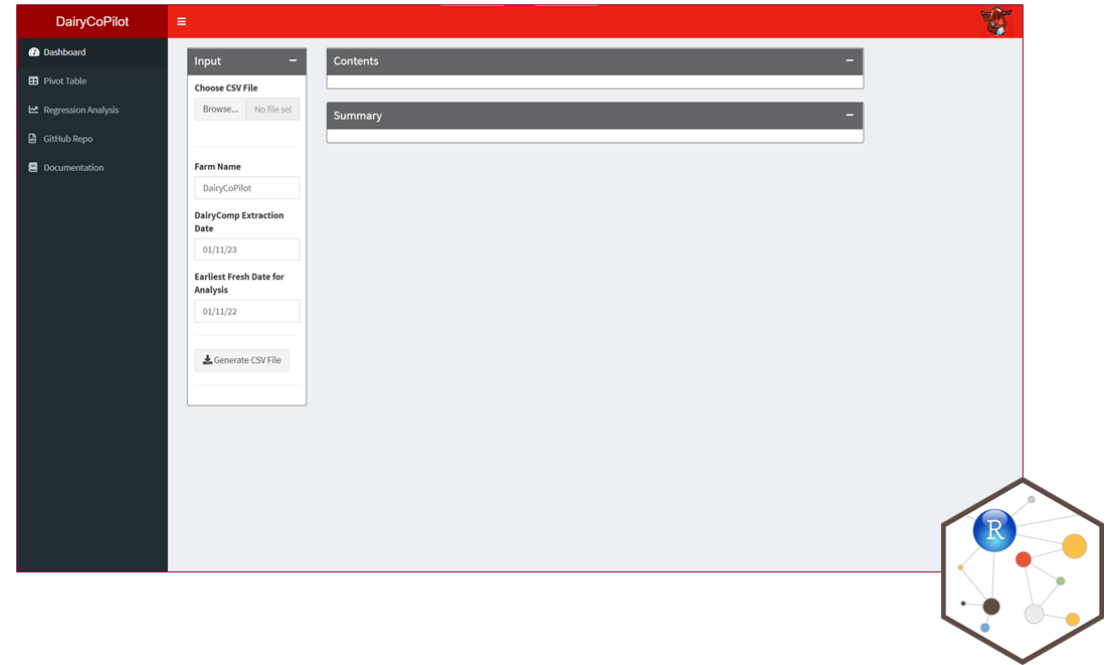
# 创建预定义变量
data$case_name <- data$date
data$case_date <- data$event
data$case_remark <- data$remark

# 转换为适当的变量类型
data$case_name <- as.character(data$case_name)
data$case_date <- as.Date(data$case_date)
```



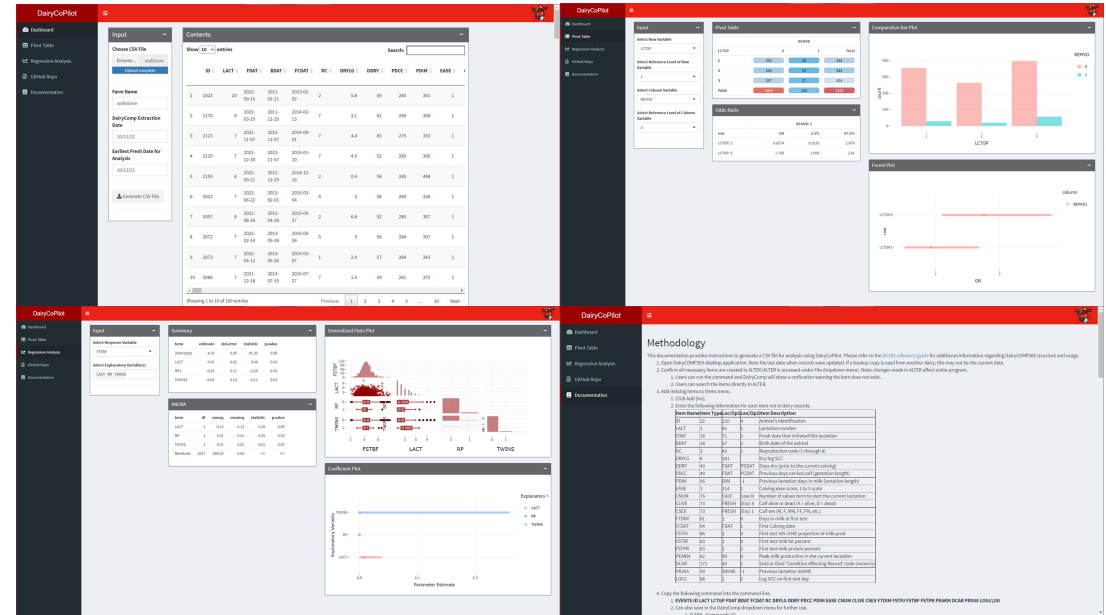
# RStudio Connect Deployment

- Standalone publishing platform
  - Dashboards
  - On-demand and scheduled reports
  - Interactive applications
- User-friendly web application
  - Access controls
  - Performance settings
  - View logs



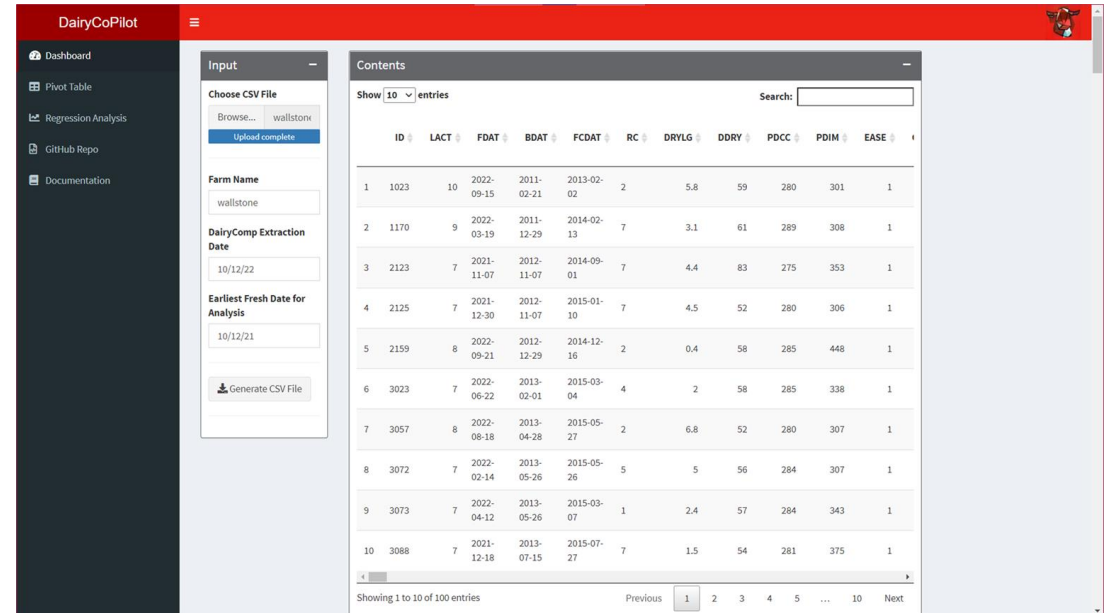
# Results

1. Input Tab
2. Pivot Table Tab
3. Regression Analysis Tab
4. Documentation Tab



# Input Tab

- Upload CSV file from DairyComp
- Download cleaned dataset to local drive
- Descriptive statistics
  - Show top entries
  - Data summary by variable type

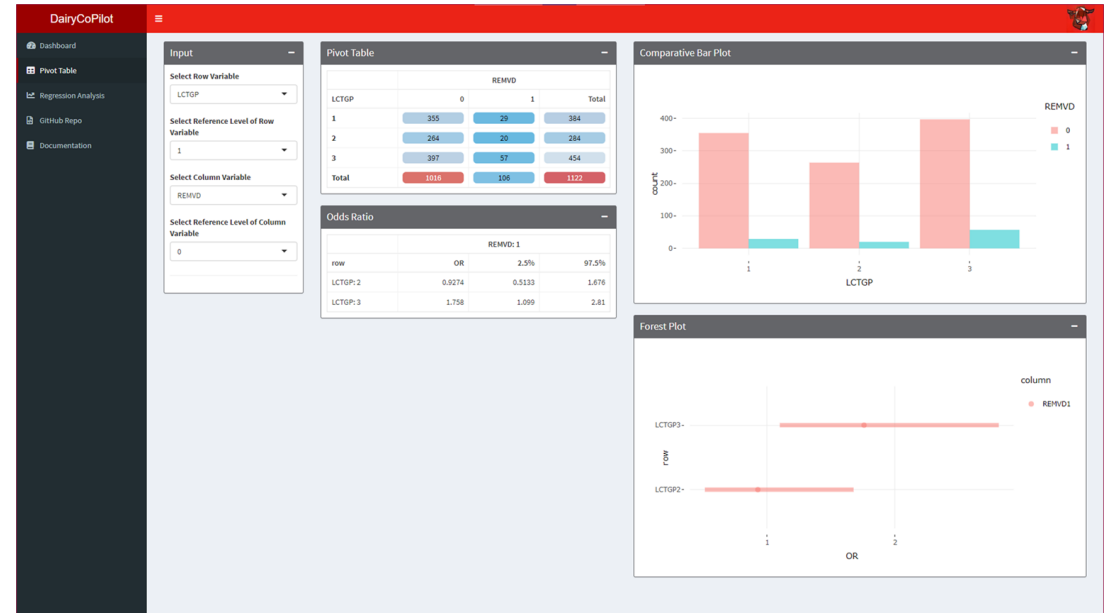


The screenshot shows the DairyCoPilot web application interface. On the left is a dark sidebar with navigation links: Dashboard, Pivot Table, Regression Analysis, GitHub Repo, and Documentation. The main content area is divided into two panels. The 'Input' panel on the left contains a 'Choose CSV File' section with 'Browse...' and 'wallstone' buttons, an 'Upload complete' button, and input fields for 'Farm Name' (wallstone), 'DairyComp Extraction Date' (10/12/22), and 'Earliest Fresh Date for Analysis' (10/12/21), followed by a 'Generate CSV File' button. The 'Contents' panel on the right displays a table of 10 entries. The table has columns: ID, LACT, FDATE, BDAT, FCDAT, RC, DRYLG, DDDRY, PDCC, PDIM, and EASE. The data rows show various farm IDs and associated dates and metrics. At the bottom of the table, it says 'Showing 1 to 10 of 100 entries' with pagination controls for Previous, 1, 2, 3, 4, 5, ..., 10, and Next.

	ID	LACT	FDATE	BDAT	FCDAT	RC	DRYLG	DDRY	PDCC	PDIM	EASE
1	1023	10	2022-09-15	2011-02-21	2013-02-02	2	5.8	59	280	301	1
2	1170	9	2022-03-19	2011-12-29	2014-02-13	7	3.1	61	289	308	1
3	2123	7	2021-11-07	2012-11-07	2014-09-01	7	4.4	83	275	353	1
4	2125	7	2021-12-30	2012-11-07	2015-01-10	7	4.5	52	280	306	1
5	2159	8	2022-09-21	2012-12-29	2014-12-16	2	0.4	58	285	448	1
6	3023	7	2022-06-22	2013-02-01	2015-03-04	4	2	58	285	338	1
7	3057	8	2022-08-18	2013-04-28	2015-05-27	2	6.8	52	280	307	1
8	3072	7	2022-02-14	2013-05-26	2015-05-26	5	5	56	284	307	1
9	3073	7	2022-04-12	2013-09-26	2015-03-07	1	2.4	57	284	343	1
10	3088	7	2021-12-18	2013-07-15	2015-07-27	7	1.5	54	281	375	1

# Pivot Table Tab

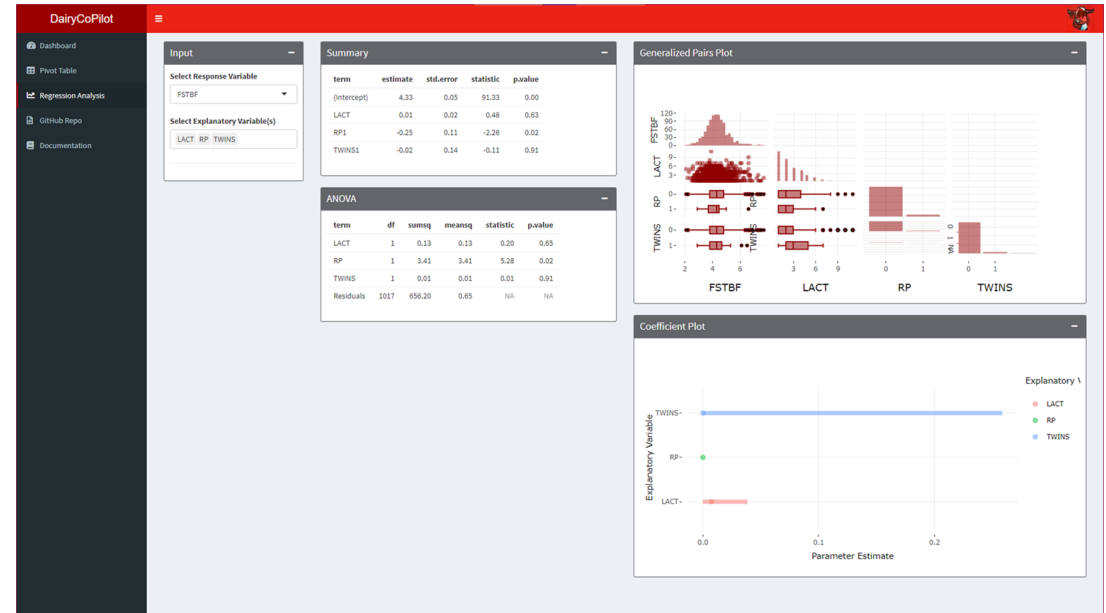
- Categorical data analysis
  - Select row and column variable
  - Select reference level for each variable
- Statistical analysis
  - Pivot table
  - Odds ratios and 95% confidence intervals
- Data visualization
  - Comparative bar plot
  - Forest plot





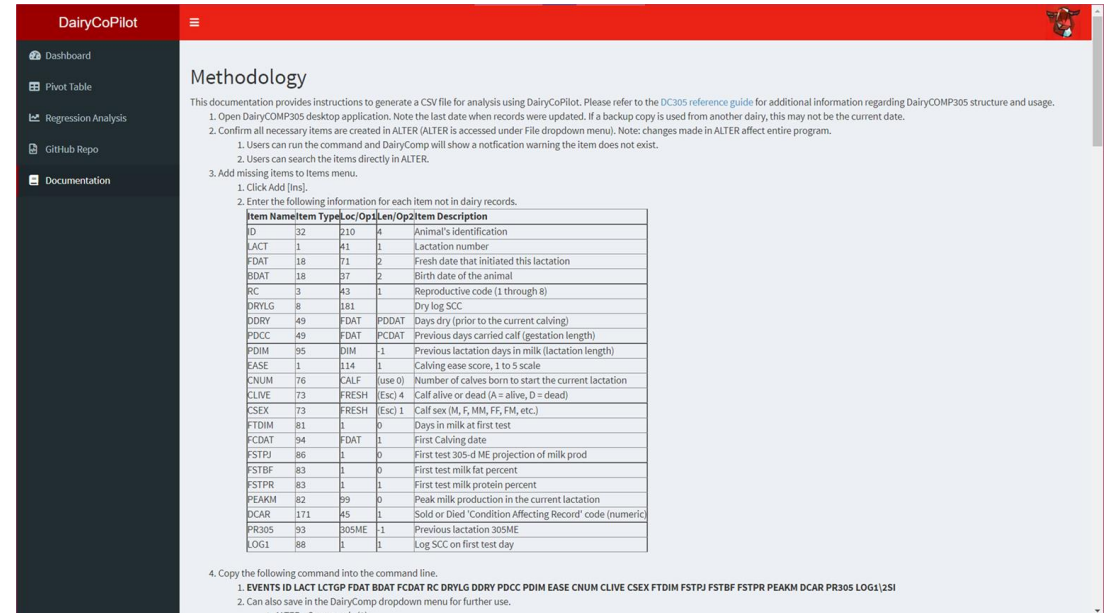
# Regression Analysis Tab

- Linear and logistic regression analysis
  - Select a response variable
  - Select multiple explanatory variables
- Statistical analysis
  - Summary table
  - ANOVA table
- Data visualization
  - Generalized pairs plot
  - Coefficient plot



# Documentation Tab

- Methodology
  - DairyComp Data Extraction
- Glossary
  - DairyCoPilot Data Cleaning
- Contacts
- Examples
  - Example Datasets
  - Instructional Videos (Upcoming)



The screenshot shows the DairyCoPilot web application interface. On the left is a dark sidebar with navigation links: Dashboard, Pivot Table, Regression Analysis, GitHub Repo, and Documentation (which is highlighted). The main content area is titled 'Methodology' and contains instructions for generating a CSV file for analysis. It includes a table of item codes and descriptions, and a list of steps for using the application.

**DairyCoPilot**

Dashboard  
Pivot Table  
Regression Analysis  
GitHub Repo  
Documentation

### Methodology

This documentation provides instructions to generate a CSV file for analysis using DairyCoPilot. Please refer to the [DC305 reference guide](#) for additional information regarding DairyCOMP305 structure and usage.

1. Open DairyCOMP305 desktop application. Note the last date when records were updated. If a backup copy is used from another dairy, this may not be the current date.
2. Confirm all necessary items are created in ALTER (ALTER is accessed under File dropdown menu). Note: changes made in ALTER affect entire program.
  1. Users can run the command and DairyComp will show a notification warning the item does not exist.
  2. Users can search the items directly in ALTER.
3. Add missing items to Items menu.
  1. Click Add [Ins].
  2. Enter the following information for each item not in dairy records.

Item Name	Item Type	Loc	Op1	Len	Op2	Item Description
ID	32	210	4			Animal's identification
LACT	1	41	1			Lactation number
FDAT	18	71	2			Fresh date that initiated this lactation
BDAT	18	27	2			Birth date of the animal
RC	3	43	1			Reproductive code (1 through 8)
DRYLG	8	181				Dry log SCC
DDRY	49	FDAT	PDDAT			Days dry (prior to the current calving)
PDCC	49	FDAT	PCDAT			Previous days carried calf (gestation length)
PDIM	95	DIM	-1			Previous lactation days in milk (lactation length)
EASE	1	114	1			Calving ease score, 1 to 5 scale
CNUM	76	CALF	(use 0)			Number of calves born to start the current lactation
CLIVE	73	FRESH	(Esc) 4			Calf alive or dead (A = alive, D = dead)
CSEX	73	FRESH	(Esc) 1			Calf sex (M, F, MM, FF, FM, etc.)
FTDIM	81	1	0			Days in milk at first test
FCDAT	94	FDAT	1			First Calving date
FSTPJ	86	1	0			First test 305-d ME projection of milk prod
FSTBF	83	1	0			First test milk fat percent
FSTPR	83	1	1			First test milk protein percent
PEAKM	82	89	0			Peak milk production in the current lactation
DCAR	171	45	1			Sold or Died "Condition Affecting Record" code (numeric)
PR305	93	305ME	-1			Previous lactation 305ME
LOG1	88	1	1			Log SCC on first test day

4. Copy the following command into the command line.
  1. `EVENTS ID LACT LCTGP FDAT BDAT FCDAT RC DRYLG DRY PDCC PDIM EASE CNUM CLIVE CSEX FTDIM FSTPJ FSTBF FSTPR PEAKM DCAR PR305 LOG1;251`
  2. Can also save in the DairyComp dropdown menu for further use.

# Case Study

- Wall-Stone dataset
  - 1000-cow dairy in Wisconsin, USA
  - 43012 events collected in 2022

<https://drive.google.com/file/d/11QAM1TTe8YizGLIshxIQsKIFCCOrfjSI/view?usp=sharing>



<https://data-viz.it.wisc.edu/dairy-copilot/>



# Summary and Future Directions

- DairyComp is still the backbone of dairy cattle management
- Tools such as DairyCoPilot expand limited analysis options
  - Increase available data
  - Decrease preprocessing time
- Upcoming features
  - Create custom columns
  - Edit factor variables
  - Edit incorrect data directly in application
- Build new features requested by users



SCAN ME

# Thank you!

