

PLS 120: Applied Statistics in Agricultural Sciences

R Statistical Playground



Explore, Experiment, and Discover

Mohammadreza Narimani
Department of Biological and Agricultural Engineering
University of California, Davis

mnarimani@ucdavis.edu

November 2024

Contents

1 Important Links	2
2 Welcome to Your Statistical Laboratory	2
3 Getting Started	2
3.1 Launching Your Playground	2
3.2 No Installation Required	3
4 What You Can Explore	3
4.1 Statistical Concepts	3
4.2 Research Applications	3
5 Your Statistical Toolkit	4
5.1 Pre-Installed Packages	4
5.1.1 Data Manipulation & Import	4
5.1.2 Visualization	4
5.1.3 Statistical Analysis	4
5.1.4 Hypothesis Testing & Modeling	4
6 Exploration Ideas	4
6.1 Beginner Explorations	5
6.2 Intermediate Challenges	5
6.3 Advanced Projects	6
7 Learning Through Experimentation	6
7.1 The Playground Philosophy	6
7.2 Effective Exploration Strategies	7
8 Bringing Your Own Data	7
8.1 Data Import Options	7
8.2 Data Upload Process	7
9 Saving Your Discoveries	7
9.1 Important: Binder is Temporary!	8
9.2 Export Options	8
10 Research Ideas by Field	8
10.1 Agricultural Sciences	8
10.2 Environmental Sciences	9
11 Statistical Concepts to Explore	9
11.1 Fundamental Concepts	9
11.2 Advanced Topics	10
12 Collaboration and Sharing	10
12.1 Learning Together	10
12.2 Building Your Portfolio	11
13 Troubleshooting and Tips	11
13.1 Common Issues and Solutions	11
13.2 Best Practices	12
14 Beyond the Playground	12
14.1 Continuing Your Statistical Journey	12
14.2 Career Applications	13

15 Getting Help	13
15.1 Support Resources	13
15.2 Contact Information	13
16 Final Thoughts	14

Important Links

Essential Course Resources

Course Website

All course materials available at:

[Course Website Link](#)

R Statistical Playground

Launch your exploration environment:

[Playground Binder Link](#)

Welcome to Your Statistical Laboratory

Welcome to the **R Statistical Playground** - your personal space for statistical exploration and discovery! This interactive environment is designed for students who want to go beyond the structured weekly assignments and dive deeper into the fascinating world of statistics.

What is the R Statistical Playground?

The playground is a cloud-based R environment where you can:

- **Experiment** with statistical concepts freely
- **Explore** your own datasets
- **Test** new ideas and hypotheses
- **Learn** through hands-on practice
- **Discover** insights in agricultural data

Getting Started

Launching Your Playground

1. Click the **Playground Binder Link** above
2. Wait 2-5 minutes for the environment to load
3. Open `R_Statistical_Playground.ipynb`
4. Start exploring!

No Installation Required

Everything you need is pre-installed and ready to use:

- **R 4.3** - Latest statistical computing environment
- **40+ Packages** - All essential statistical libraries
- **Jupyter Interface** - Interactive notebook environment
- **Sample Data** - Built-in datasets for practice

What You Can Explore

Statistical Concepts

Experiment with Core Concepts:

- **Descriptive Statistics** - Means, medians, standard deviations
- **Probability Distributions** - Normal, t, chi-square distributions
- **Hypothesis Testing** - t-tests, ANOVA, chi-square tests
- **Confidence Intervals** - Understanding uncertainty
- **Regression Analysis** - Linear and multiple regression
- **Correlation Analysis** - Relationships between variables
- **Data Visualization** - Creating compelling plots
- **Sampling Methods** - Bootstrap and simulation techniques

Research Applications

Real-World Applications to Explore:

- **Agricultural Experiments** - Crop yield analysis, treatment comparisons
- **Environmental Studies** - Climate data, biodiversity assessments
- **Quality Control** - Process monitoring, defect analysis
- **Survey Research** - Opinion polls, market research
- **Biological Research** - Growth studies, genetic analysis
- **Economic Analysis** - Price trends, market behavior
- **Social Sciences** - Behavioral studies, demographic analysis
- **Engineering** - Reliability testing, performance optimization

Your Statistical Toolkit

Pre-Installed Packages

All these powerful R packages are ready for your exploration:

5.1.1 Data Manipulation & Import

Data Handling:

`dplyr` - Data manipulation and transformation
`tidyr` - Data tidying and reshaping
`readr` - Fast CSV file reading
`readxl` - Excel file import
`data.table` - High-performance data operations

5.1.2 Visualization

Creating Beautiful Plots:

`ggplot2` - Grammar of graphics for elegant visualizations
`plotly` - Interactive plots and dashboards
`corrplot` - Correlation matrix visualizations
`pheatmap` - Heatmaps for complex data patterns

5.1.3 Statistical Analysis

Advanced Statistics:

`car` - Companion to Applied Regression
`agricolae` - Agricultural experimental design
`multcomp` - Multiple comparisons procedures
`emmeans` - Estimated marginal means
`broom` - Tidy statistical output

5.1.4 Hypothesis Testing & Modeling

Statistical Testing:

`lmtest` - Linear model diagnostic tests
`nortest` - Normality testing procedures
`lawstat` - Robust statistical methods
`MASS` - Modern Applied Statistics functions

Exploration Ideas

Beginner Explorations

Start with These Ideas:

1. Built-in Dataset Analysis

- Explore the `iris` dataset - flower measurements
- Analyze `mtcars` - automobile performance data
- Investigate `diamonds` - gemstone characteristics

2. Statistical Concept Testing

- Generate random data and test normality
- Compare different sampling methods
- Visualize the Central Limit Theorem

3. Visualization Experiments

- Create different plot types for the same data
- Experiment with color schemes and themes
- Build interactive plots with `plotly`

Intermediate Challenges

Level Up Your Skills:

1. Hypothesis Testing Practice

- Design and conduct your own t-tests
- Compare different statistical test approaches
- Explore Type I and Type II errors

2. Regression Modeling

- Build predictive models with multiple variables
- Compare linear vs. polynomial regression
- Assess model assumptions and diagnostics

3. Data Simulation

- Generate synthetic agricultural datasets
- Simulate experimental designs
- Test statistical power under different scenarios

Advanced Projects

Push Your Boundaries:

1. Research Question Investigation

- Formulate your own research hypotheses
- Design appropriate statistical analyses
- Interpret results in scientific context

2. Method Comparison Studies

- Compare different statistical approaches
- Evaluate robustness under various conditions
- Document advantages and limitations

3. Real Data Analysis

- Import your own datasets
- Conduct comprehensive statistical analysis
- Create publication-quality reports

Learning Through Experimentation

The Playground Philosophy

Embrace the Learning Process:

- **Make Mistakes** - Errors are learning opportunities
- **Ask "What If?"** - Test your statistical intuition
- **Compare Methods** - Try different approaches to the same problem
- **Visualize Everything** - Plots reveal hidden patterns
- **Document Discoveries** - Keep track of interesting findings
- **Share Insights** - Discuss findings with classmates
- **Stay Curious** - Follow interesting tangents
- **Think Critically** - Question assumptions and results

Effective Exploration Strategies

Maximize Your Learning:

1. **Start Simple** - Begin with basic concepts before advancing
2. **Build Gradually** - Add complexity step by step
3. **Test Assumptions** - Verify what you think you know
4. **Seek Patterns** - Look for consistent relationships
5. **Challenge Results** - Ask if findings make sense
6. **Connect Concepts** - Link new learning to previous knowledge
7. **Apply Context** - Consider real-world implications
8. **Reflect Often** - Think about what you've learned

Bringing Your Own Data

Data Import Options

The playground supports various data formats:

Supported File Types:

- **CSV Files** - `read.csv("filename.csv")`
- **Excel Files** - `read_excel("filename.xlsx")`
- **Text Files** - `read.table("filename.txt")`
- **R Data Files** - `load("filename.RData")`
- **JSON Files** - `fromJSON("filename.json")`

Data Upload Process

1. Click the **Upload** button in the left panel
2. Select your data file from your computer
3. Wait for upload to complete
4. Use appropriate R function to load data
5. Begin your analysis!

Saving Your Discoveries

Important: Binder is Temporary!**Remember to Save Your Work:**

- Binder sessions are temporary and will be deleted
- Always download your notebook before closing
- Save interesting code snippets for future use
- Export visualizations and results
- Document important findings

Export Options**Ways to Save Your Work:**

1. **Download Notebook** - File → Download as → Notebook (.ipynb)
2. **Export HTML** - File → Download as → HTML (.html)
3. **Save Plots** - Right-click plots to save images
4. **Copy Code** - Select and copy useful code snippets
5. **Export Data** - Save processed datasets for future use

Research Ideas by Field**Agricultural Sciences****Agricultural Research Questions:**

- How do different fertilizer treatments affect crop yield?
- What factors influence soil pH across different regions?
- How does weather variability impact harvest timing?
- What's the relationship between plant density and productivity?
- How do organic vs. conventional methods compare?
- What factors predict pest outbreak severity?
- How does irrigation timing affect water use efficiency?
- What variables influence post-harvest quality?

Environmental Sciences

Environmental Research Questions:

- How has local temperature changed over decades?
- What factors influence biodiversity in different habitats?
- How do pollution levels vary across urban areas?
- What predicts species abundance in ecosystems?
- How do conservation efforts impact wildlife populations?
- What factors influence water quality in watersheds?
- How does land use change affect carbon storage?
- What variables predict ecosystem resilience?

Statistical Concepts to Explore

Fundamental Concepts

Core Statistical Ideas:

- **Central Tendency** - Mean, median, mode behavior
- **Variability** - Standard deviation, variance, range
- **Distribution Shapes** - Normal, skewed, bimodal patterns
- **Outliers** - Detection methods and impact assessment
- **Correlation** - Strength and direction of relationships
- **Causation** - Distinguishing from correlation
- **Sampling** - Random vs. systematic approaches
- **Bias** - Sources and mitigation strategies

Advanced Topics

Deeper Statistical Exploration:

- **Power Analysis** - Sample size determination
- **Effect Sizes** - Practical vs. statistical significance
- **Multiple Testing** - Controlling Type I error rates
- **Non-parametric Methods** - Robust alternatives
- **Bootstrap Methods** - Resampling techniques
- **Model Selection** - Choosing appropriate analyses
- **Assumption Checking** - Validating test requirements
- **Interaction Effects** - Complex variable relationships

Collaboration and Sharing

Learning Together

Collaborative Learning Opportunities:

- Share interesting findings with classmates
- Discuss different approaches to the same problem
- Compare results from different analytical methods
- Help each other troubleshoot coding challenges
- Exchange ideas for research questions
- Peer review each other's analyses
- Create study groups for advanced topics
- Present discoveries to the class

Building Your Portfolio

Document Your Journey:

- Keep a collection of interesting analyses
- Save well-commented code examples
- Create a library of useful functions
- Document lessons learned from mistakes
- Build a portfolio of visualizations
- Record insights about statistical methods
- Maintain a list of research questions
- Track your skill development progress

Troubleshooting and Tips

Common Issues and Solutions

Troubleshooting Guide:

- **Binder Won't Load** - Try refreshing or clearing browser cache
- **Package Errors** - Use `install.packages("package_name")`
- **Data Import Issues** - Check file format and path
- **Memory Problems** - Restart kernel or reduce data size
- **Plot Not Showing** - Ensure proper syntax and data
- **Code Errors** - Check for typos and missing parentheses
- **Slow Performance** - Simplify analysis or restart session
- **Lost Work** - Always save frequently!

Best Practices

Maximize Your Success:

- Start each session by loading required libraries
- Comment your code to remember your thinking
- Use meaningful variable names
- Save your work frequently
- Test code with small examples first
- Visualize data before analyzing
- Check assumptions before applying tests
- Interpret results in context

Beyond the Playground

Continuing Your Statistical Journey

Next Steps in Your Learning:

- **Advanced Courses** - Multivariate statistics, experimental design
- **Specialized Software** - SAS, SPSS, Python for data science
- **Research Projects** - Apply skills to real research questions
- **Internships** - Gain practical experience in data analysis
- **Conferences** - Present findings at scientific meetings
- **Publications** - Contribute to peer-reviewed literature
- **Professional Development** - Statistical consulting, data science
- **Lifelong Learning** - Stay current with new methods

Career Applications

Statistical Skills in Your Future:

- **Research Scientist** - Design and analyze experiments
- **Data Analyst** - Extract insights from complex datasets
- **Quality Control** - Monitor and improve processes
- **Consultant** - Provide statistical expertise to organizations
- **Policy Analyst** - Inform decisions with data
- **Agricultural Extension** - Translate research for practitioners
- **Environmental Monitoring** - Track ecosystem changes
- **Business Intelligence** - Support strategic planning

Getting Help

Support Resources

When You Need Assistance:

- **Course Materials** - Review weekly tutorials and examples
- **R Documentation** - Use `?function_name` for help
- **Online Resources** - Stack Overflow, R-bloggers, RStudio guides
- **Textbooks** - Statistical methods and R programming guides
- **Video Tutorials** - YouTube channels for R and statistics
- **Study Groups** - Collaborate with classmates
- **Office Hours** - Meet with instructor or TA
- **Online Forums** - R community support groups

Contact Information

Mohammadreza Narimani

Email: mnarimani@ucdavis.edu

Department of Biological and Agricultural Engineering, UC Davis

Office Hours: Thursdays 10 AM - 12 PM (Zoom)

Zoom Link: Join Office Hours

Final Thoughts

The R Statistical Playground is more than just a computing environment - it's your gateway to statistical thinking and data-driven discovery. Every great statistician started with curiosity and a willingness to experiment. Use this space to:

Embrace Your Statistical Journey:

- **Be Curious** - Ask questions about everything you see
- **Stay Patient** - Learning statistics takes time and practice
- **Think Critically** - Always question your results and methods
- **Enjoy Discovery** - Celebrate insights and "aha!" moments
- **Share Knowledge** - Teach others what you learn
- **Keep Exploring** - There's always more to discover
- **Apply Learning** - Use statistics to solve real problems
- **Never Stop Learning** - Statistics is a lifelong journey

Remember: The best way to learn statistics is by doing statistics. This playground gives you the freedom to explore, experiment, and discover at your own pace. Make mistakes, ask questions, and most importantly - have fun with data!