Pesticide-Driven Bee Mortality: *An Introduction to Survival Analysis in R*

MEGAN BLACK/ VIRGINIA COMMONWEALTH UNIVERSITY 2022

# Learning Objectives

## Environmental Science

Students will learn about the ecological effects of pesticide application and how these chemicals affect pollinator species, specifically bumble bees.

## Data Analysis in R Studio

Students will receive an introductory lesson to survival analysis and how to conduct this in R Studio. Students will create survivability models, plot survival curves, and run chi-square tests to interpret the models’ statistical significance.

## Conservation

After reviewing the results of the focal paper and the conclusions from their own data analyses, students will be able to evaluate the effects of pesticides on bee populations.

# Prerequisites

**Students should have both R and RStudio installed on their computers, with a working knowledge of the software’s layout.**

Introduction

# The Environmental Issue: Pesticides and Bee Decline

Bees are pollinators, which means that they have significant ecological importance. However, populations of bees and other pollinator species are experiencing considerable declines across the world. Scientists, like those who write our lesson’s focal paper, are concerned that pesticide use may be a major driver of bee loss.

Pesticides include a wide range of chemicals and are often marketed as directed towards a certain target species. Categories of pesticides include insecticides, herbicides, fungicides, and rodenticides. Unfortunately, the idea of a “target species” is little more than a false comfort to consumers. Pesticides have the ability to spread beyond the location of their application, seeping into the soil and waterways. They can have toxic effects on any organism they interact with. Many plants and animals (even humans!) respond to pesticide exposure through the bioaccumulation of toxins: this means that the chemicals enter their systems and are stored in cells, where they sometimes even increase in magnitude. Most bees, however, typically experience mortality. Affected bees can be found in transitionary landscapes that have been exposed to some sort of pesticide application: they crawl around dazed and confused, unable to retract their protruding tongues, until they die.

Chemical pesticides can be organized into different families. The organochlorine hydrocarbons, now banned in the United States, included popular chemicals such as DDT. Use of these pesticides was halted following the publication of Rachel Carson’s *Silent Spring*, which brought attention to their abilities to biomagnify and accumulate in both ecosystems and individuals. Organic hydrocarbons were largely replaced with organophosphates and carbamates, which alter nerve function and lead to weakness or paralysis in exposed organisms. Organophosphates have still been shown to bioaccumulate in exposed organisms.

Not included in these groups is glyphosate, the active ingredient in many herbicides. Glyphosate is a phosphonic acid that attacks enzymes in plant systems, and is used as a broad-spectrum herbicide. It is generally regarded as safe to humans and animals, but there is an ongoing debate about the chemical’s toxicity. Additionally, pesticides can include other chemical components without having to identify them on product labels. These **surfactant co-formulants** have the potential to be deadlier to non-target organisms than the identified active ingredients. Regulatory agencies consider these ingredients to be “inert” and do not require testing to ensure their safety.

Pesticides are associated with both industrial and commercial use. They are used in the agricultural industry where they are applied to crops in order to deter pests and increase yield. Pesticides are also available for consumer purchase, and are applied to suburban lawns, golf courses, and other privately-owned landscapes by individuals. Insecticides are used to deter “pest species” such as mosquitos, and herbicides are applied to kill “weeds” that threaten the growth of nonnative turfgrass species. While studies have been conducted to assess impacts of insecticide application on bee populations, there is a gap in the literature regarding herbicides. Do these chemicals also have deadly effects on pollinators? If so, are the active ingredients to blame?

# The Focal Paper: “Roundup causes high levels of mortality following contact exposure in bumble bees”

Published by Straw, Carpentier, & Brown, 2021 in the Journal of Applied Ecology

## Understanding the Focal Paper

The authors of this paper are concerned with pollinator declines and worried that pesticides could be to blame. Herbicides are the most widely applied pesticide group and have broadly considered to be “bee safe”-- regulatory bodies even allow for direct application on foraging bees. Roundup was chosen as the herbicide for this study because it is the world’s most popular herbicide brand.

Straw et al. hypothesize that glyphosates are not the drivers of bee mortality associated with herbicide application and exposure. Instead, they argue that co-formulants other than glyphosates are to blame.

The aim of this study was to determine the mortality rates of bumble bees associated with different herbicide treatments. Ten commercial bumble bee populations were used and over fifty bees were exposed to treatment within each experiment. Five total experiments were conducted:

1. Are the impacts of consumer and agricultural Roundup products comparable?
2. Does mortality still occur with a 1:1 dilution of consumer Roundup?
3. Does mortality still occur with a 1:3 dilution of consumer Roundup?
4. Does an alternative GHB (Weedol) cause mortality?
5. Does the Roundup formulation without glyphosate cause mortality?

## The Data

For this tutorial, we will be utilizing data from experiments one, four, and five. First, we will look at whether the impacts of consumer and agricultural Roundup are comparable. This will be provided as an example within the code and lesson for students to reference. After walking through each line of the provided example code, students will run the same models and tests for experiments four and five, thus recreating Figures 1-3 in the focal paper. This will help us to determine the toxicity of Roundup and whether its active ingredient, glyphosate, is related to bee mortality.

Survival Analysis

# What is Survival Analysis?

Survival analysis is the analysis of time-to-event data, describing the length of time from a time origin to an endpoint. Essentially, a variety of statistical approaches allow us to determine the time it takes for an event of interest to occur. Survival analyses are used in fields such as engineering and sociology, but one of the best examples of this approach is its incorporation within clinical trials and cancer studies to analyze the survival time of patients.

Depending on the type of survival analysis being performed, different types of events are recorded. These include relapse, progression, and death. In our focal paper’s study, the event recorded was death. The period from origin to the occurrence of the event of interest (death) is called survival time. Here, our time origin is the application of herbicides.

The **survival probability**, or survival function S(t), is the probability that an individual survives from the time origin to a future specified time, t. This can be calculated using the Kaplan-Meier method, a non-parametric method used to estimate the survival probability from observed survival times. The survival probability at time ti, S(ti), is calculated as followed:

S(ti)= S(ti–1)(1– di/ni)

Where:

* S(ti-1) = the probability of being alive at ti-1
* Ni = the number of patients alive just before ti
* Di = the number of events at ti
* T0 = 0, S(0) = 1

## Chi-Square Tests

In the focal paper, the authors ran chi-square tests to determine whether there were significant differences between the treatments for each experiment. Chi-square tests measure independence and can show us if there is a relationship between two categorical variables. This works by comparing the observed frequencies to the expected frequencies if there is truly no relationship between the two variables. Here, we can assume both a null and an alternative hypothesis:

* H0: The two variables are independent and have no relationship. The distribution of the outcome is independent of these groups.
* H1: The two variables relate to each other. There is a difference between the distribution of responses to the outcome variable among the comparison groups.

The output of a chi-square test shows us three things: the test statistic (X-squared), the degrees of freedom (df), and the p-value. The test statistic measures the difference between the observed and expected frequencies of the outcomes, essentially telling us how well a model compares to actual, observed data. The larger this value is, the greater the probability is of a significant difference. Degrees of freedom help us to determine whether the null hypothesis can be rejected, and the p-value represents the test’s statistical significance. If our p-value is less than 0.05, the test can be considered significant and we can reject our null hypothesis. This is the general statistical practice utilized in scientific studies, including the focal paper we are using for this lesson.

## Supplementary Resources

* In-depth [article](https://www.sciencedirect.com/science/article/pii/S1756231716300639) on survival analysis as a concept
* [Tutorial](http://www.sthda.com/english/wiki/survival-analysis-basics) for coding survival analysis in R Studio
* [Information](https://www.rdocumentation.org/packages/survminer/versions/0.4.9/topics/ggsurvplot) for the ggsurvplot() function, including customizations
* [Walkthrough](https://statsandr.com/blog/chi-square-test-of-independence-in-r/#conclusion-and-interpretation) of chi-square tests in R

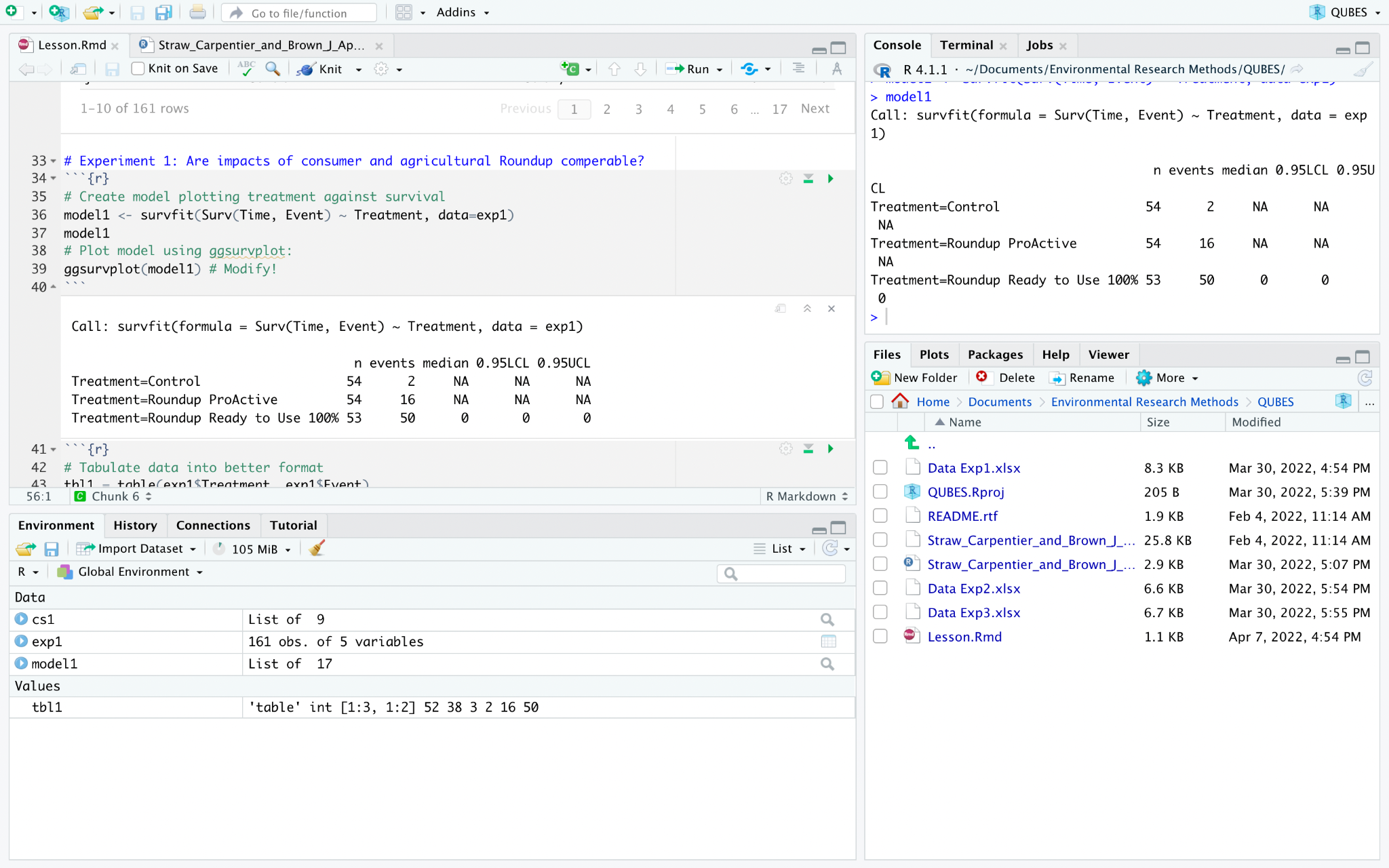
# Survival Analysis in R Studio

Now we are going to walk through the process of survival analysis in R Studio. **Students should open the lesson’s .Rmd file in R and run through the example code line by line.** This section of the lesson elaborates on the R code associated with Experiment 1. Before we can utilize the survival function, there are a few housekeeping things we need to do first:

* Open your lesson .Rmd file in R, in the same location/ project as your data files
* Install and require the packages needed for this lesson
* Read in the .csv or .xlsx for Experiment 1
* Use the colnames() and is.factor() commands to ensure data looks correct

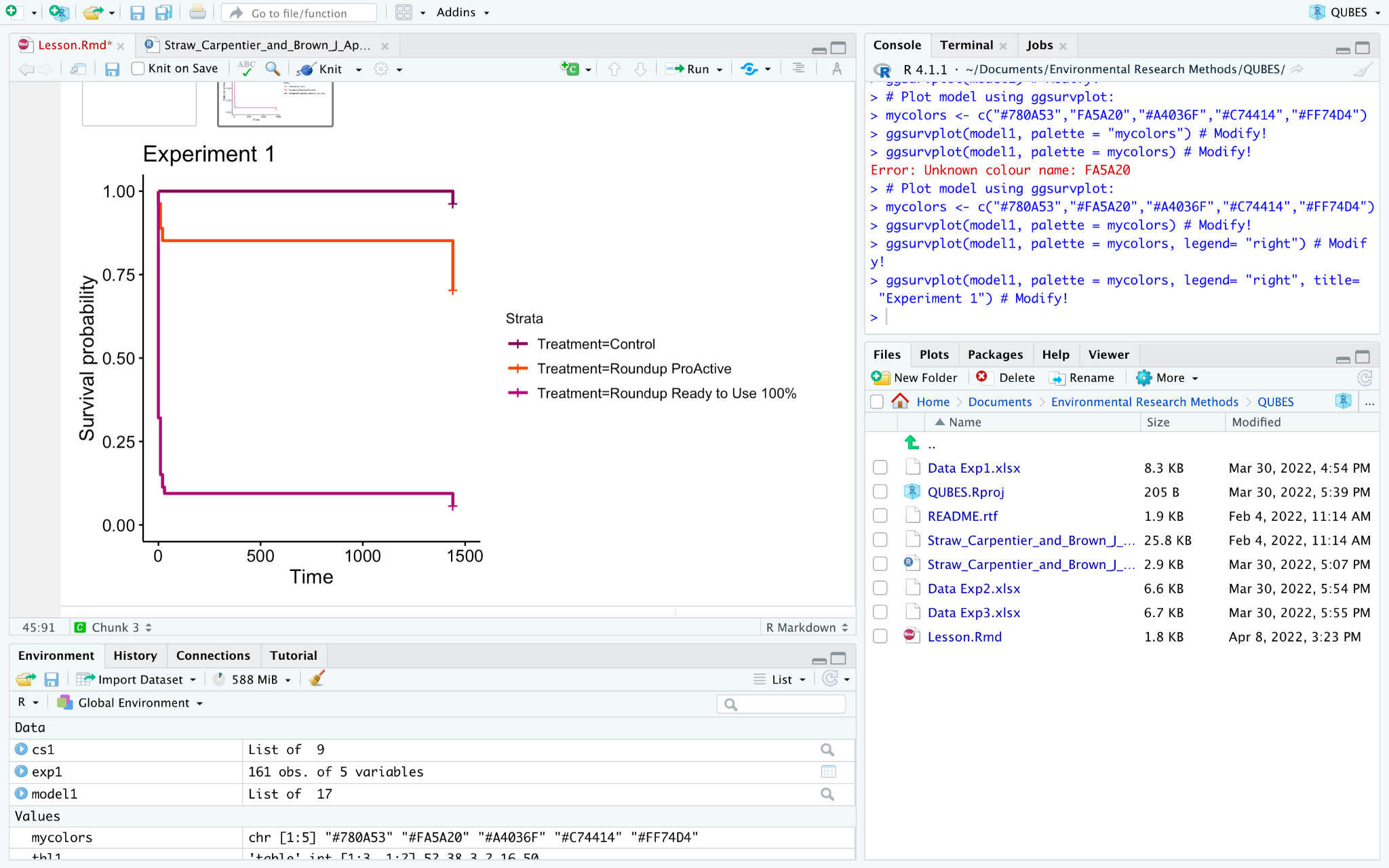
## Survival Probability Model

To model survivability, we will use the function survfit(). Survival, assessed by considering time and event, is modeled as a function of treatment. This function utilizes the Kaplan-Meier method. Calling the created model forward will display a summary of it within your .Rmd. This summary prints the number of observations, number of events, and the median survival with its confidence limits. The output for our Experiment 1 model is not summarized well by these parameters, but it is important to know what these variables are measuring for future models.



## Plotting Survival with ggsurvplot()

Next, we will use ggsurvplot() to model our survivability curves. This is a function that runs through the ggplot2 package and allows us to easily generate survival curves. The authors of the focal paper simply ran their models through this function, but we will make alterations in our code to create more aesthetically pleasing visuals than the ones included in the focal paper.

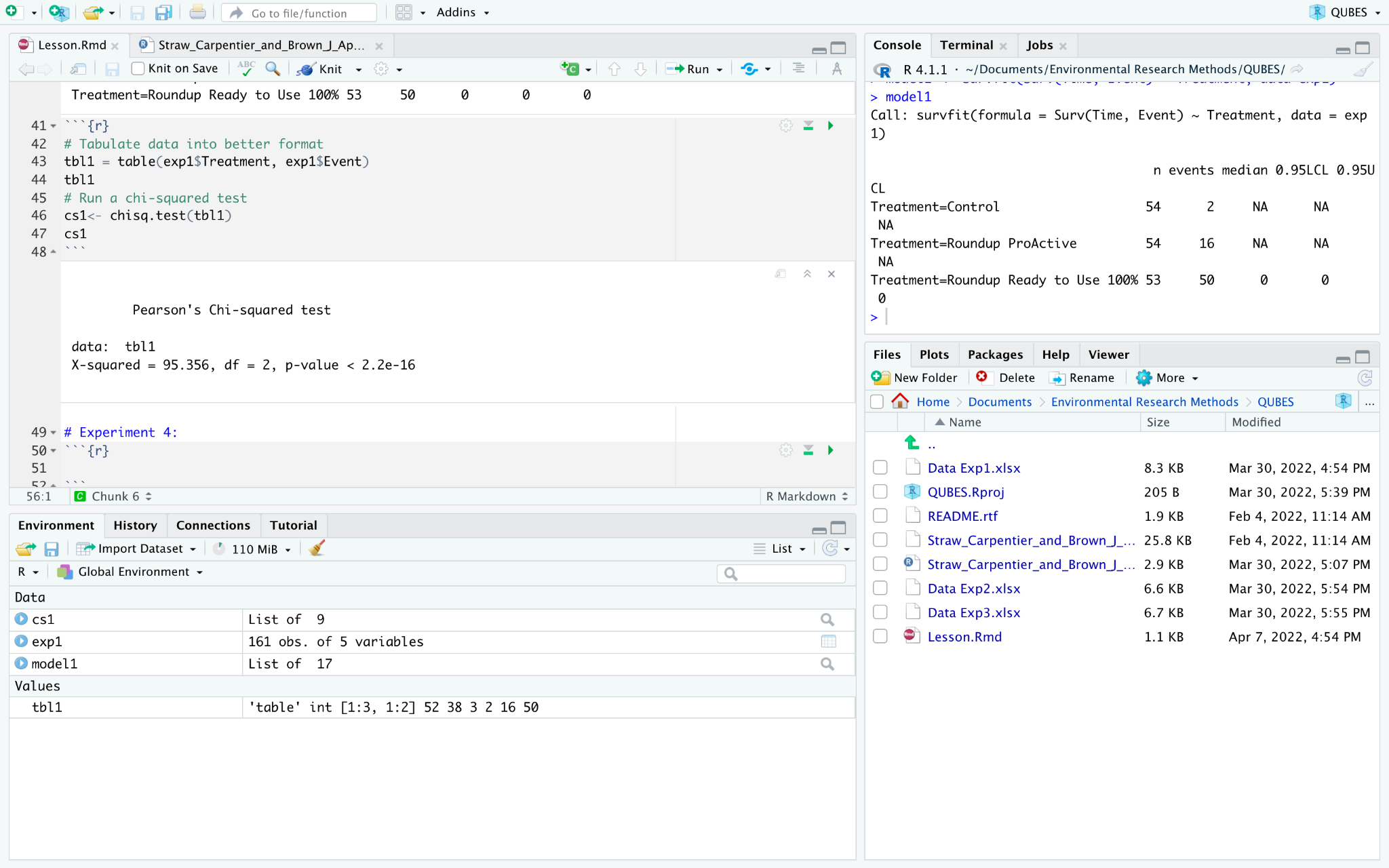


Look at this figure and try to understand what is happening. As time passed during Experiment 1, the survival probability for the various treatments are shown to be different. The control stays at 100% throughout, and the agricultural Roundup ProActive treatment hovers around 80%. The line for consumer-marketed Roundup Ready to Use drops dramatically towards the beginning of the experiment and shows a high, fast mortality response.

**BONUS**: In my code, I create my own color palette to use in my plots. You can make your own based on my code and colors that you like! The website [Coolors.com](https://coolors.co/a4036f-fa5a20-c74414-780a53-ff74d4) has a fun random palette generator that gives you hex codes to plug into your code.

## Running Chi-Square Tests

Before we run a chi-square test, we have to tabulate our data into a better format. Make a table of Experiment 1’s Treatment and Event variables. Then, plug the table into a chi-square test using the function chisq.test() and interpret its results. We have a large x-squared value, few degrees of freedom, and a very small p-value. It looks like the results from Experiment 1 were statistically significant, which allows us to reject the null hypothesis and conclude that different variables had different effects on the response. For our first experiment, this means that different types of Roundup (consumer and agricultural) had different mortality responses on our bee populations.



# On Your Own

Repeat the above process on your own for experiments four and five:

* Read in the experiment’s .csv
* Ensure that the data has been read into R correctly
* Make a survival probability model
* Plot survival probability with ggsurvplot()
* Tabulate the data and run a chi-squared test

# Test Your Knowledge

# Now that you have gone through coding Experiments 4 & 5 on your own, answer the following questions:

## Based on our findings from the first experiment, what is the difference between consumer and agricultural Roundup?

## Did your survival plot for Experiment 4 look different compared to the first experiment? How do you interpret the results displayed by this figure?

1. **How would you interpret the results of the chi-square test for Experiment 4?**
2. **Did your survival plot for Experiment 5 look different than previous ones? How do you interpret the results being displayed?**
3. **How would you interpret the results of the chi-square test for Experiment 5?**
4. **How do your results compare with those of the focal publication? Were your own findings similar or different?**
5. **With Experiments 4 & 5 in mind, what are this study’s major implications for conservation? Can you identify a “call to action” from these findings?**

# References

Kartsonaki, C. (2021, March 31). *Survival analysis*. Diagnostic Histopathology. Retrieved May 5, 2022, from https://www.sciencedirect.com/science/article/pii/S1756231716300639

Straw, E.A., Carpentier, E.N., & Brown, M.J.F. (2021). Roundup causes high levels of mortality following contact exposure in bumble bees. *Journal of Applied Ecology, 58*, 1167-1176. doi:10.1111/1365-2664.13867

*Survival analysis basics*. STHDA. (n.d.). Retrieved April 5, 2022, from http://www.sthda.com/english/wiki/survival-analysis-basics