# C-15:
# First Steps With Integrating R & RStudio Into Actuarial Workflow

Thomas Hartl, Milliman
Jean-François Greeff, PwC

## Antitrust Notice

- **The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.**
- **Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.**
- **It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.**

2

## Disclaimer

This presentation, and any code contained, is intended solely for educational purposes and presents information of a general nature. It is not intended to guide or determine any specific individual situation and persons should consult qualified professionals before taking specific actions. The views expressed in this presentation are those of the presenter, and not those of Milliman , Inc. ("Milliman"). Milliman does not intend to benefit and assumes no duty or liability to any person or entity who receives this work. This is not intended to be an actuarial opinion or actuarial advice. **Milliman makes no warranty as to the use, application or performance of the code contained in this presentation.**

## First Steps with Integrating R into Actuarial Workflow

Learning Goals

- Understanding what components are necessary to run (and edit) R scripts.
- Executing an R script interactively from within RStudio.
- Exchanging data with R by reading from, and writing to *.csv files.
- Understanding how print ready pdf exhibits can automatically be generated using R Markdown.
- Understanding how integrating simulation and exhibit generation can enhance our quality control workflow.

# Components Needed

… and what they do

## Open Source Galore!



## First Steps with Integrating R into Actuarial Workflow

Components Needed – **R Interpreter**

- Core engine for executing programs written in the R language (aka R scripts).
- Long-standing Open Source project.
- Commercial versions available for taking advantage of parallel computation and cloud hosting.
- Very popular in academia for implementing cutting edge statistical methods.
- Interpreted language, ideal for interactively analyzing data and developing models.
- Relies on functional programming. So, if your prior programming experience is with VBA, some things may take some getting used to.



## First Steps with Integrating R into Actuarial Workflow

Components Needed – **R Packages**


- Avoid reinventing the wheel: “There’s a package for that!”
- Packages are libraries, often designed to implement a specific technique, or to provide tools that facilitate efficiently working with data.
- Pick and choose: you get to select which packages to download to your computer, and which packages to activate for a particular R session.
- R can be very flexible, even to the extent that the basic syntax and evaluation rules can be changed. Learning how to get the most out of a package can therefore take some time.

## First Steps with Integrating R into Actuarial Workflow

Components Needed – **RStudio**


- You can run R scripts without RStudio ...
- … but working with a proper integrated development environment (IDE) makes things a lot easier.
- Provides debugging facilities.
- Integrates with extensions such as R Markdown, or Shiny.
- … de facto standard for now; Microsoft started offering Visual Studio Tools for R, but this is not as mature an IDE as RStudio is at this point.



## First Steps with Integrating R into Actuarial Workflow

Components Needed – **R Markdown** / **LaTeX**

- R Markdown is an extension that allows mixing R calculations with formatting instructions for various target output media, notably HTML and pdf.
- Actually, R Markdown can also incorporate output from other languages (and was inspired by Python notebooks).
- To generate multi-page pdf documents such as slide presentations or reports, R Markdown uses LaTeX as the typesetting engine.
- LaTeX is a typesetting language that has been around for decades. For the longest time, it was the only game in town for print quality mathematical formulas.
- TeX Live is a TeX/LaTeX implementation maintained by the TeX User Group.

# Hands-On Demo

… slides for reference after session

## First Steps with Integrating R into Actuarial Workflow

Getting Started

- Make sure you have installed R, RStudio, R Packages, and TeX Live (see separate instructions for installation).
- Make sure that the following files end up in C:\AnnualMeetingC15:
  - Simulation Model.R
  - Report Exhibits.rmd
  - Property Mixed Exponential.csv, GL Mixed Exponential.csv, Recall Mixed Exponential.csv

## First Steps with Integrating R into Actuarial Workflow

Running R Script Interactively

- Start RStudio and use File -> Open File … to open C:\AnnualMeetingC15\Simulation Model.R
- The “Run” icon allows you to run specific lines (based on your selection in the script pane; top left).
- The “Source” icon will run the entire active script (you may have multiple open scripts).
- The console pane (bottom left) shows you what is going on in R proper.
- The environment pane (top right) shows you the state of the R interpreter.
- The auxiliary pane (bottom right) has other useful info, such as help.



## First Steps with Integrating R into Actuarial Workflow

Reading and Writing Data

- Fine to use read.csv and write.csv for small data sets.
- For larger data sets use fread and fsave from the data.table package.
- Data from *.csv file is stored in a structure called “data frame” (columns of data vectors; each column can be a different data type, but all elements within one column must have the same data type).
- Separate file for each data.frame.
- There is also a package (look for “xlsx”) for creating *.xlsx files (or adding tabs to a *.xlsx file); this can be used to create a “link-able” interface to Excel.
- Other file formats are available, but may be harder to interpret. If you want to save a large data set of simulation results, you may have to look into this.

## First Steps with Integrating R into Actuarial Workflow

Generating PDF Exhibits

- It is possible to maintain all R code and formatting instructions in one R Markdown template, but …
- … nicely formatted tables and consistent headers are trickier than one might think. Hence, we opted to separate the logic of the R code for the simulation from what we need to do to format the output.
- Our "Simulation Model.R" script performs the simulation and then calls the "Report Exhibits.rmd" template to generate the report.
- The *.rmd template "inherits" the environment from the calling R script. So, the exhibits are generated by directly querying both the results and the internal structure of the simulation model.



# Review / Conclusion

… What have we learned?

## Name the component! Why do we need it?






## First Steps with Integrating R into Actuarial Workflow

Tasks Performed

- Read in data for model specification from *.csv files.
- Ran simulation.
- Summarized simulation results.
- Saved results (separately for detail and summary) to *.csv files.
- Created pdf exhibits that summarize both the simulation results and the model assumptions that went into the simulation.
- Performance compares favorably with @Risk: simulation with 100,000 iterations takes about 20 seconds (@Risk will take 5 to 8 minutes + the time you need to manually cut and paste the results and print the exhibits); a simulation with 1,000,000 iterations takes less than 3 minutes.

## First Steps with Integrating R into Actuarial Workflow

Potential Enhancements to Workflow

- Once an R script and R Markup template have been set up, the cycle of changing model parameters and re-running the simulation is highly efficient.
- Since the same script does the simulation, summarizes the results, and generates the exhibits, the risk of failing to keep results and documentation in sync is largely eliminated.
- Since the exhibit updating process is automated, it becomes feasible to create exhibits that aid both peer-review and checking of the simulation model by colleagues who are not familiar with R programming.



## First Steps with Integrating R into Actuarial Workflow

What comes next?

- There are many “Getting started with R” tutorials available. A nice offering comes from datacamp.com: “Introduction to R” (this course is free, but they charge for other content).
- Can you see yourself using an R based simulation model like the one presented today for your work?
- What are the obstacles you perceive to using something like the demonstrated set-up for your work?
- Feel free to reach out to the presenters if you have further questions:
  - Thomas Hartl <thomas.hartl@milliman.com>
  - Jean-François Greeff <jean.francois.greeff@pwc.com>