



```
course = "Improving your statistical inferences through simulation studies in R"
```

```
lesson_iteration = 4  
lesson_title = "Tidy data and tidy simulations"
```

```
auth = "Ian Hussey"  
dept = "Psychology of Digitalisation"
```

Foundational Concepts

in Monte Carlo Simulation Studies

Core components of a simulation

1. Generate pseudo-random data set with known properties
2. Analyse data with a statistical method
3. Repeat 1 & 2 many times ('iterations')
4. Summarize results across iterations
5. Make it an experiment
 - Systematically vary parameters in Step 1 (between factor)
 - Compare different ways to do Step 2 (within factor)



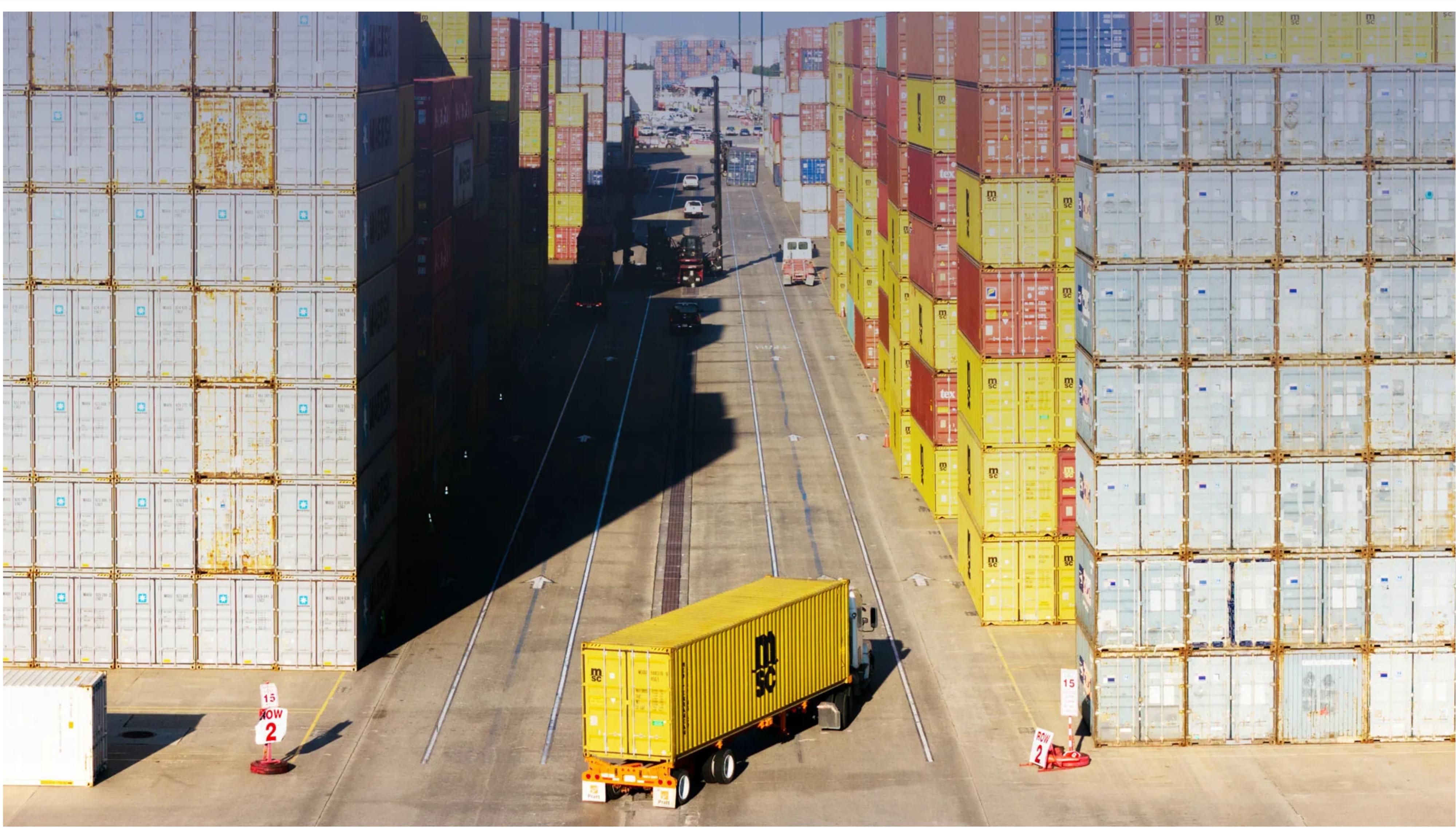
aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"



aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"



aut = "Ian Hussey";

dept = "Psychology of Digitalisation || Digitalisation of Psychology"

Treat the simulation code workflow like an **assembly line**

Every stage of the process **takes tidy inputs** and **produces tidy outputs**



Treat the simulation code workflow like an **assembly line**

Every stage of the process **takes tidy inputs and produces tidy outputs**

This forces you to clarify your thinking and make individual parts of the simulation work

(Steps inside functions can be less tidy)



Treat the simulation code workflow like an **assembly line**

Every stage of the process **takes tidy inputs and produces tidy outputs**

Most errors / getting stuck with simulations involve:

- Mixing the different steps together in the same function
- Not making inputs and outputs tidy



tidy data, tidy workflow

tidy workflow

maximising transparency and reusability

tidy inputs, tidy outputs

maximise code & workflow reusability

maximise transparency of intermediate steps

A tibble: 1 × 10									
estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	method	alternative
0.01230826	0.08646291	0.07415465	0.08756086	0.9303142	198	-0.2648943	0.2895108	Two Sample t-test	two.sided

tidy workflow

maximising transparency and reusability

tidy data, tidy workflow

How do I know if its ‘tidy’ or not?

- Data frames or tibbles, including nested ones (tabular data)
- One-column-one-variable
- One-row-one-observation
- Read Wickham’s ‘Tidy Data’

If you’re using **lists**, its not tidy.

If you’re using **S4 objects**, it’s probably not tidy.

Foundational Concepts

in Monte Carlo Simulation Studies

Example of Different Ways to Combine Factors in the Design of a Simulation Study.

