Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

# Part 1: Data

This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Abstract

Select 50 different virtual hosts and get revenue and engagement data for their last 50 live streams. The size of such a data set is sufficient for us to perform some simple analysis, but it is not too large, which makes the analysis process too cumbersome. Most of these 50 virtual anchors come from the VirtualReal project (VR is true), but some of them come from other unions or individuals. Their names, projects or unions they belong to, and user IDs are all manually entered into a csv file for subsequent data grabbing.

## Availability

Data **are** publicly available

Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

### Publicly available data

Data are available online at: https://github.com/wandleshen/vup-data-analysis

Data are available as part of the paper’s supplementary material.

Data are publicly available by request, following the process described here:

Data are or will be made available through some other mechanism, described here:

### Non-publicly available data

Discussion of lack of publicly available data:

## Description

### File format(s)

CSV or other plain text: csv

Software-specific binary format (.Rda, Python pickle, etc.):

Standardized binary format (e.g., netCDF, HDF5, etc.): h5

Other (described here):

### Data dictionary

Provided by the authors in the following file(s):

Data file(s) is (are) self-describiing (e.g., netCDF files)

Available at the following URL:

https://github.com/wandleshen/vup-data-analysis

### Additional information (optional)

See data details at https://github.com/wandleshen/vup-data-analysis/blob/master/data-analysis-tutorial.ipynb

# Part 2: Code

## Abstract

I use do data analysis with Rmarkdown. First read data with python virtual environment in ananconda. (You maybe need to modify this part by your environment.) Then use R to do other things: EDA, random forest, regression adjustment, causal forest and matching method with both regression adjustment and causal forest.

I will show both regression adjustment and causal forest with(out) matching method. The results are quite similar.

## Description

### Code format(s)

Script files

R  Python  Matlab

Other:

Package

R  Python  MATLAB toolbox

Other:

Reproducible report

R Markdown  Jupyter notebook

Other:

Shell script

Other (described here):

### **Supporting software requirements**

Version of primary software used

Python version 3.9.13; R version 4.2.1

Libraries and dependencies used by the code

Python package: pandas version 1.4.4

R packages: reticulate version 1.28, GGalary version 2.1.2, pheatmap version 2.1.2, ggplot2 version 3.4.1, grf version 2.2.1, lubridate version 1.9.2, marginaleffects version 0.11.1, randomForest version 4.7-1.1, reshape2 version 1.4.4, MatchIt version 4.5.3, dyplr version 2.3.1, extrafont version 0.11

### Supporting system/hardware requirements (optional)

Macbook M1 Air with MacOS 12.6.3

### Parallelization used

No parallel code used

Multi-core parallelization on a single machine/node

Number of cores used:

Multi-machine/multi-node parallelization

Number of nodes and cores used:

### License

MIT License (default)

BSD

GPL v3.0

Creative Commons

Other (described here):

### Additional information (optional)

https://github.com/jiahui-xin/vup-causal-inference

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

Any numbers provided in text in the paper

The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

All tables and figures in the paper

Selected tables and figures in the paper, as explained and justified here:

## Workflow details

### Location

The workflow is available:

As part of the paper’s supplementary material

In this Git repository: https://github.com/jiahui-xin/vup-causal-inference

Other:

### Format(s)

Single master code file

Wrapper (shell) script(s)

Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

Text file (e.g., a readme-style file) that documents workflow

Makefile

Other (more detail in 'Instructions' below)

### Instructions

Just run the vup.Rmd. You will see vup.htmland the results accompanied with code will be contained.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

<1 minute

1-10 minutes

10-60 minutes

1-8 hours

>8 hours

Not feasible to run on a desktop machine, as described here:

### Additional documentation (optional)

# Notes (optional)