



Visualization of incomplete and imputed data

With the R package **ggmice**

H. I. Oberman

Using **mice** for missing data

The R package **mice** has become the de facto standard to handle the ubiquitous problem of incomplete data. With **mice**, missing data points are ‘imputed’ (i.e. filled in) to obtain several completed data sets. Filling in the missing data multiple times allows for a valid representation of the uncertainty due to missingness.¹

Visualizations for **mice** with **ggplot2**

Enhance a **mice** imputation workflow with **ggmice** for visualizations of incomplete and/or imputed data. The resulting **ggplot** objects may be easily manipulated or extended. Use **ggmice** to inspect missing data, develop imputation models, evaluate algorithmic convergence, or compare observed versus imputed data.²

Get started with **ggmice**

The **ggmice** package can be installed from CRAN with `install.packages('ggmice')`. Subsequently, the packages vignettes may be retrieved with `??ggmice`. More extensive documentation can be found on the package website: amices.org/ggmice. Package contributions and bug reports are welcome through the development repository.²

Code chunk 1

Set-up the environment with packages for imputation and visualization.

```
library(mice)
library(ggplot2)
library(ggmice)
```

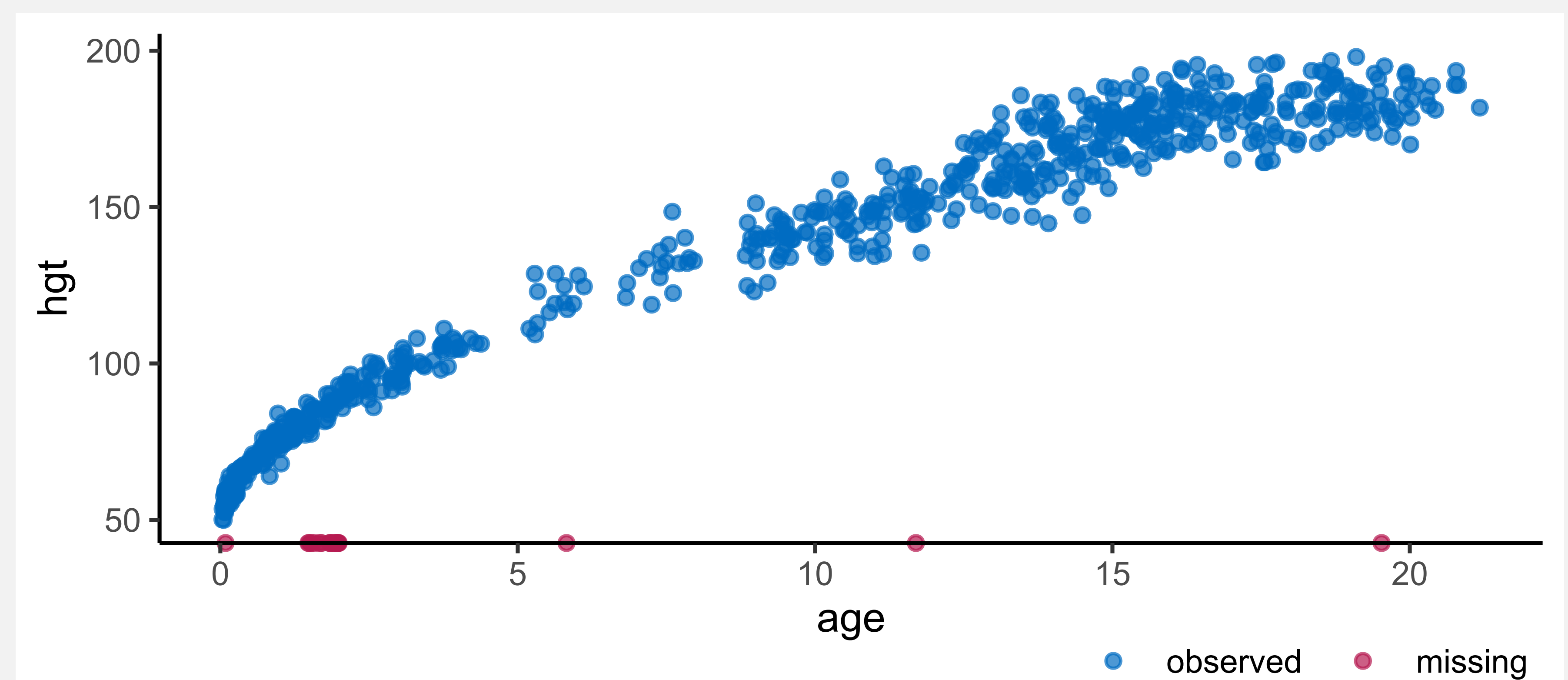
Code chunk 2

Visualize the incomplete boys data set from the **mice** package with **ggmice**.

```
ggmice(
  boys,
  aes(age, hgt)) +
  geom_point()
```

Figure 1

Scatterplot of height (in cm) by age (in years) in the incomplete **boys** data set. Missing values in the height variable are visualized with **ggmice** on the x-axis.



Code chunk 3

Impute the missing values in the **boys** data set with **mice**.

```
imp <- mice(
  boys,
  m = 1)
```

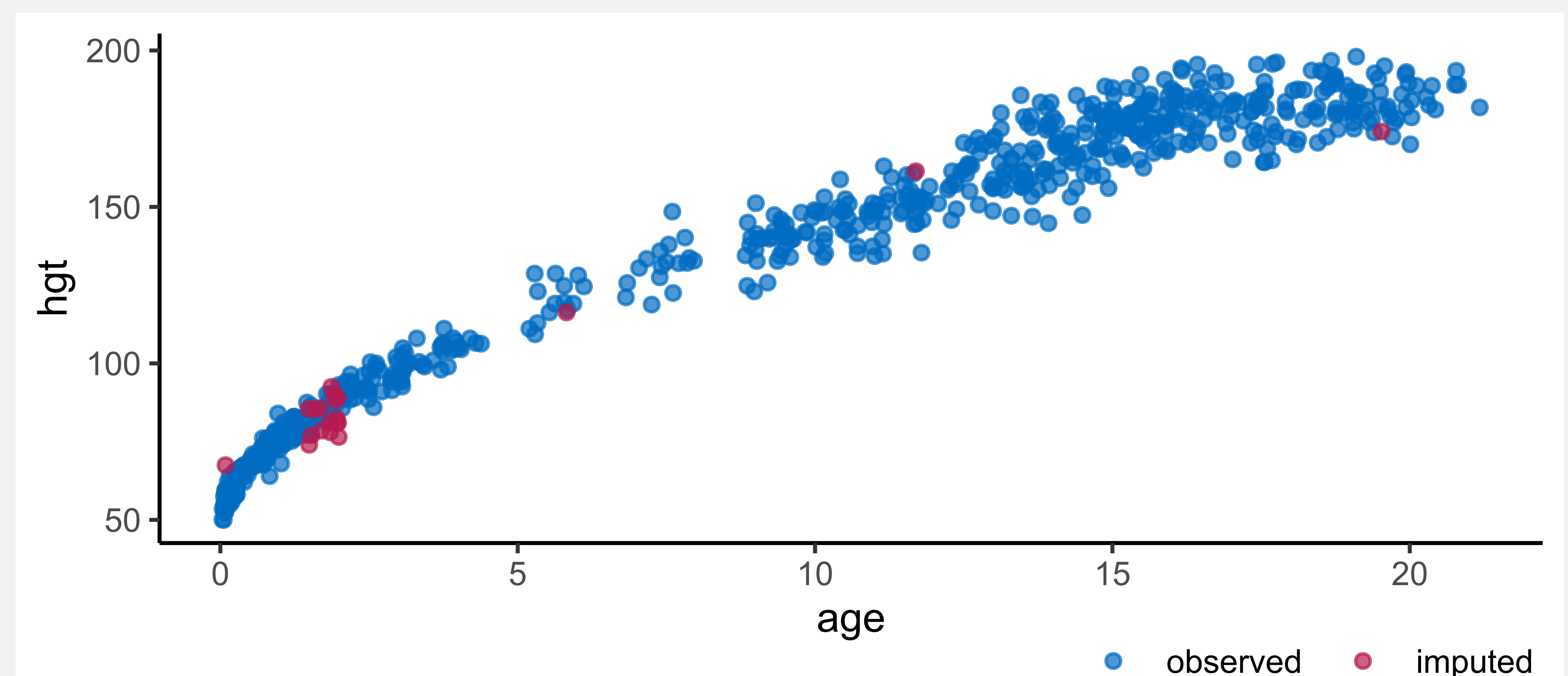
Code chunk 4

Visualize the imputed boys data set with **ggmice**.

```
ggmice(
  imp,
  aes(age, hgt)) +
  geom_point()
```

Figure 2

Scatterplot of height (in cm) by age (in years) in the imputed **boys** data set. Missing values in the height variable are filled in with **mice** and visualized with **ggmice**.



References

1. Van Buuren, S. (2018). Flexible imputation of missing data. Chapman and Hall/CRC. stefvanbuuren.name/fim/
2. Oberman, H. I. (2022). **ggmice**: Visualizations for “mice” with “ggplot2”. doi.org/10.5281/zenodo.6532702

Acknowledgements

The **ggmice** package is developed with guidance and feedback from Gerko Vink, Stef van Buuren, and others. This project has received funding from the European Union's Horizon 2020 research and innovation programme under ReCoDID grant agreement No 825746.

