



# TREES

Paula Boks, Midproject, December 2024



# Data



## Content

- 13 columns and 71,069 rows
- Type and measurements,  
location, alley, sponsorship
- Categorical and numerical

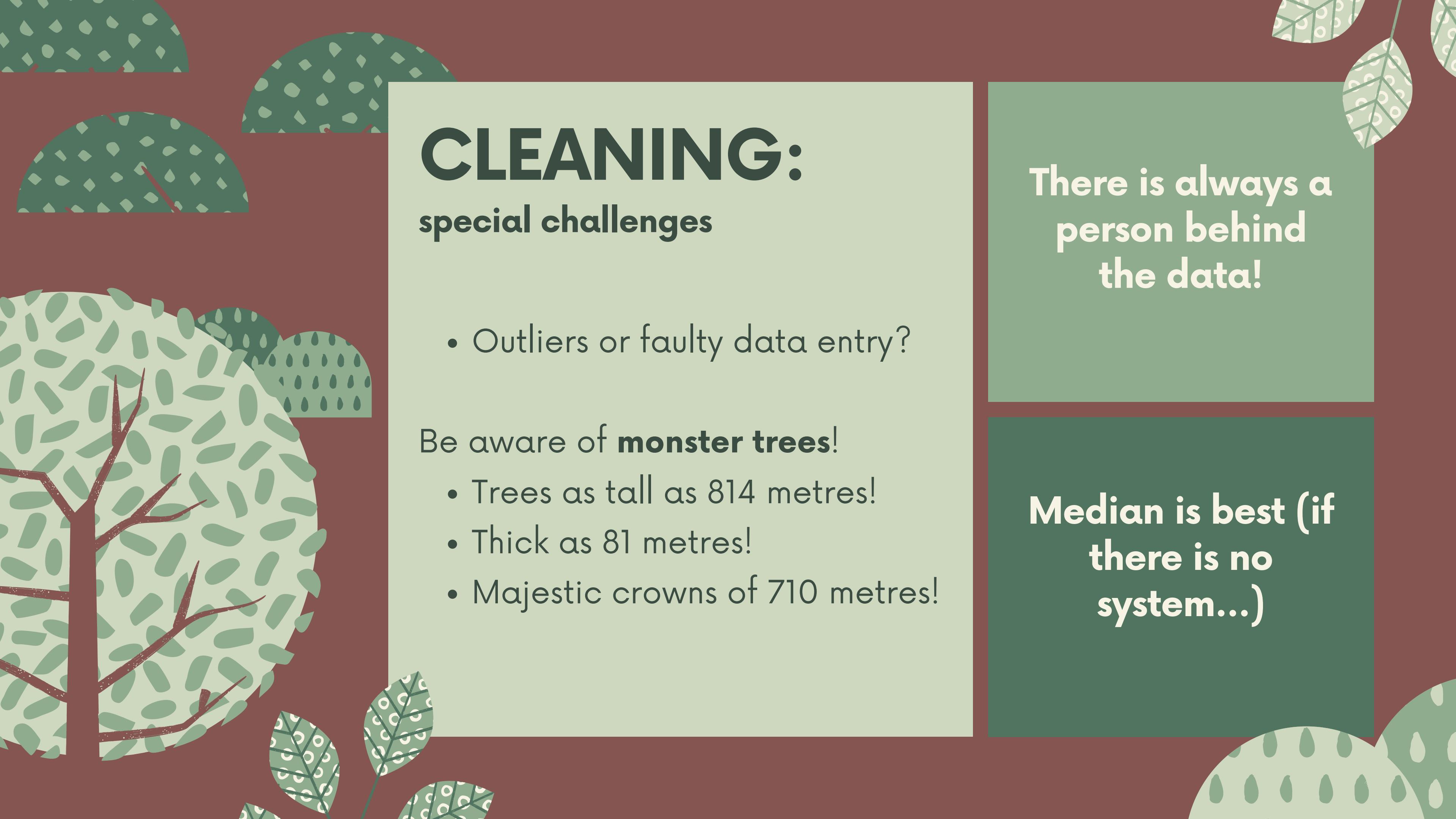
## First Steps

- Translating to English
- Limiting the columns

# Research Question and Hypotheses

What kind of trees are there, what characteristics do they have and how are these related?

1. Some tree types occur much more often.
2. Some tree types are more often part of an alley.
3. Some tree types are more likely to be sponsored.



# CLEANING: special challenges

- Outliers or faulty data entry?

Be aware of **monster trees**!

- Trees as tall as 814 metres!
- Thick as 81 metres!
- Majestic crowns of 710 metres!

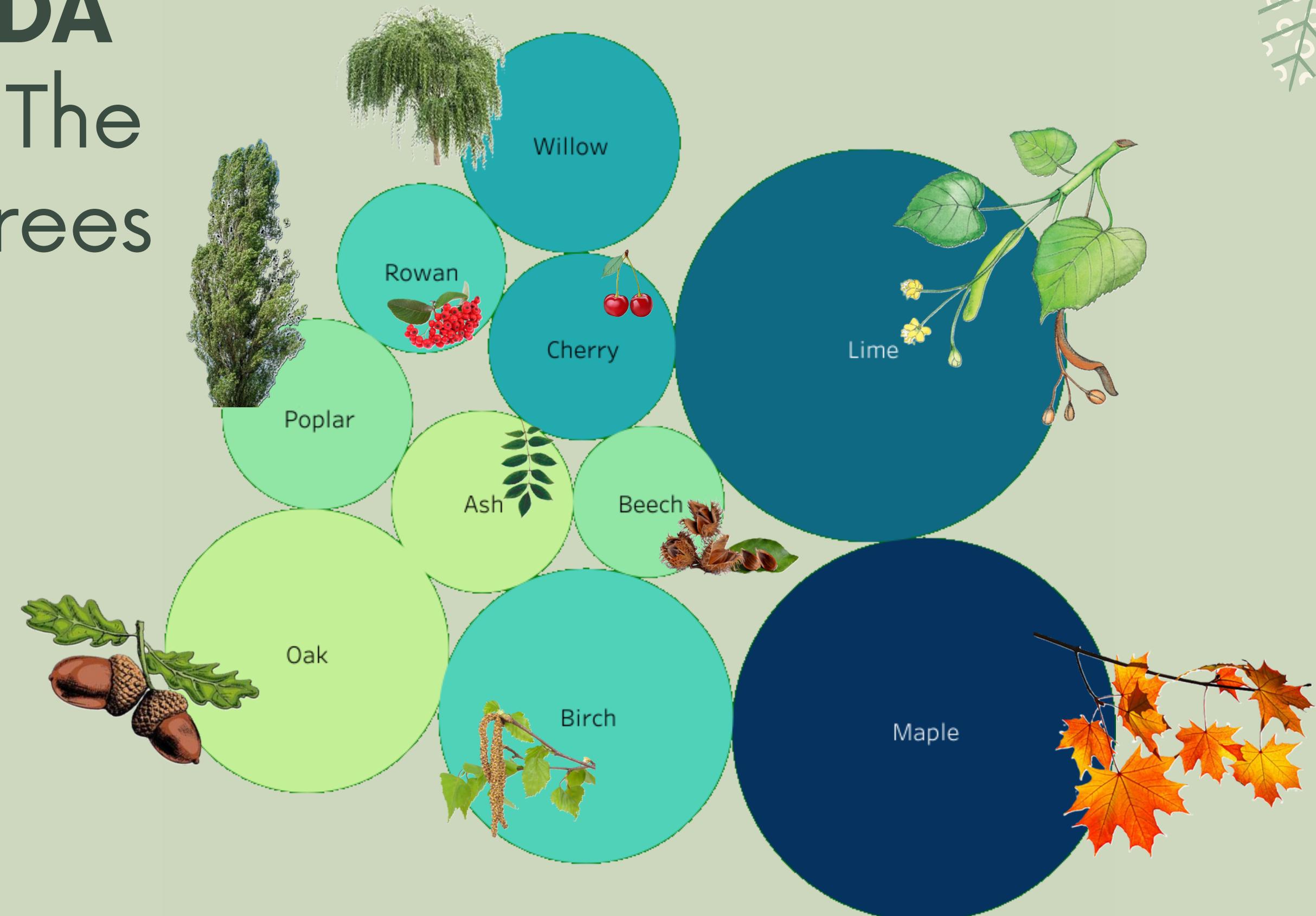
There is always a  
person behind  
the data!

Median is best (if  
there is no  
system...)



# EDA - The Trees

Number of Trees by Genus



# EDA - Numericals only

## Tree Measurements:

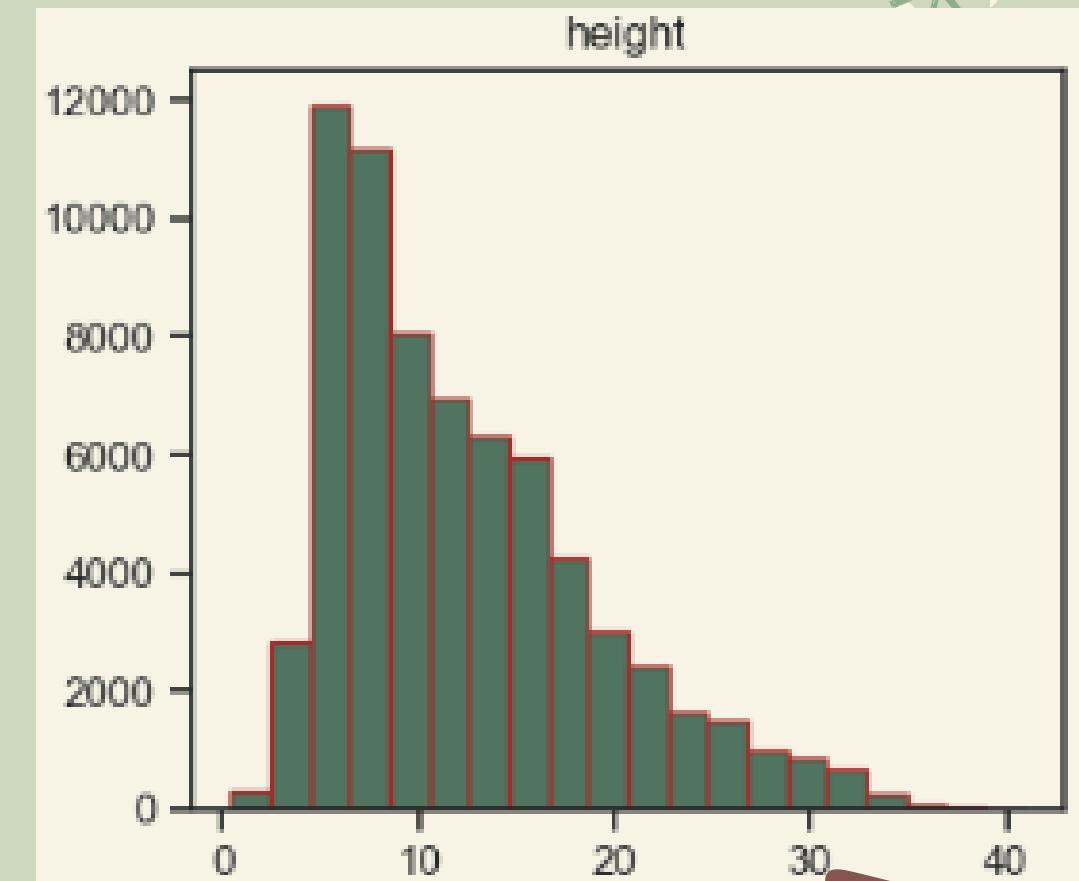
- Most trees are in the lower values area, some are much bigger ("baby trees")
- Height and trunk circumference are strongly correlated (linear & monotonic)

## Spearman coefficient (monotonic)



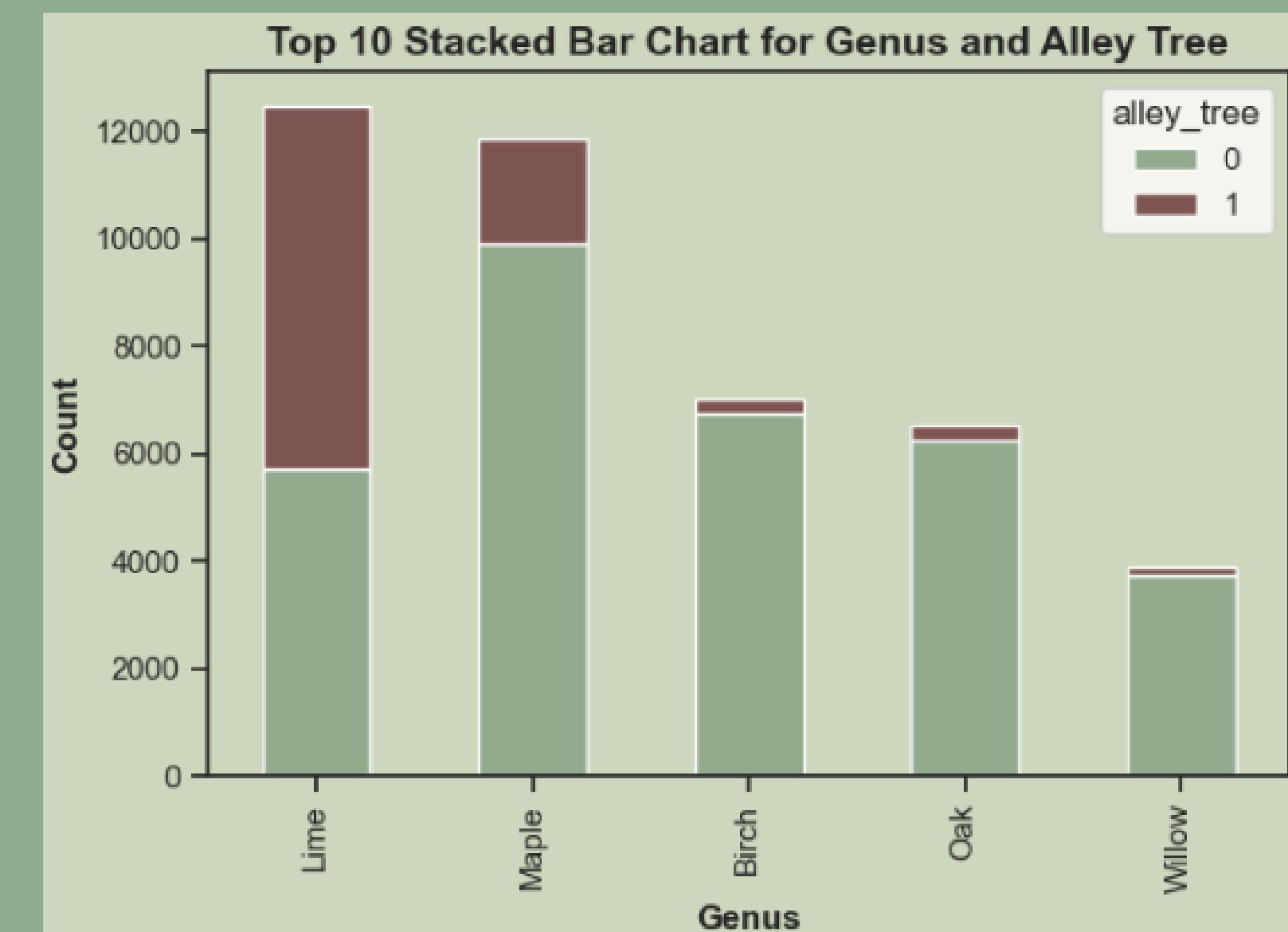
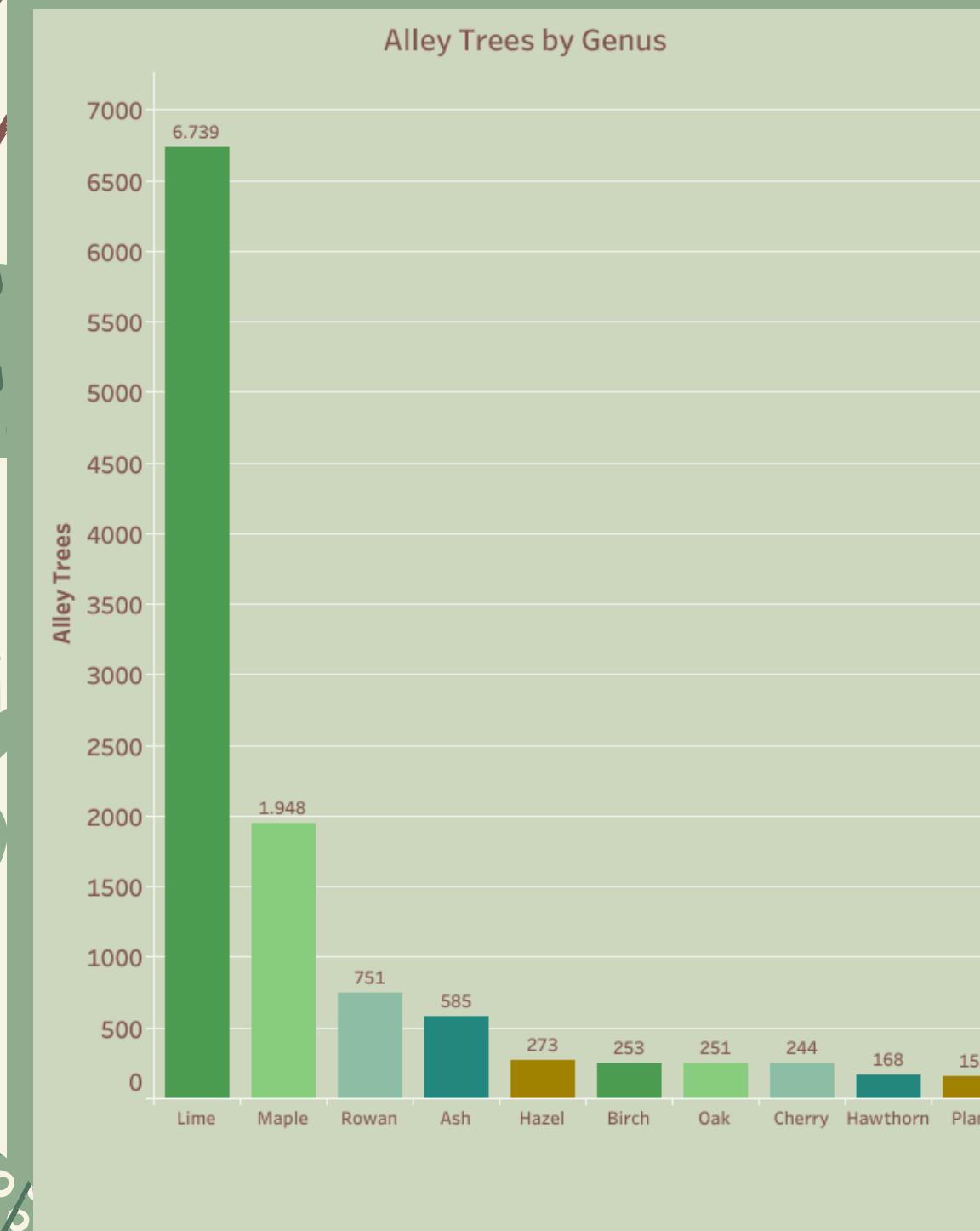
	height	trunk_circumference	crown_diameter
height	1.00	0.78	0.18
trunk_circumference	0.78	1.00	0.21
crown_diameter	0.18	0.21	1.00

Histogram tree heights



# EDA - Genus vs. Alley

If a tree is part of an alley **highly correlates** with its genus



# EDA - Genus vs. Measures

**Distribution** and **mean** values:

- The four genera of trees behave very differently regarding height, trunk and crown

*Oak trees*

**mean**

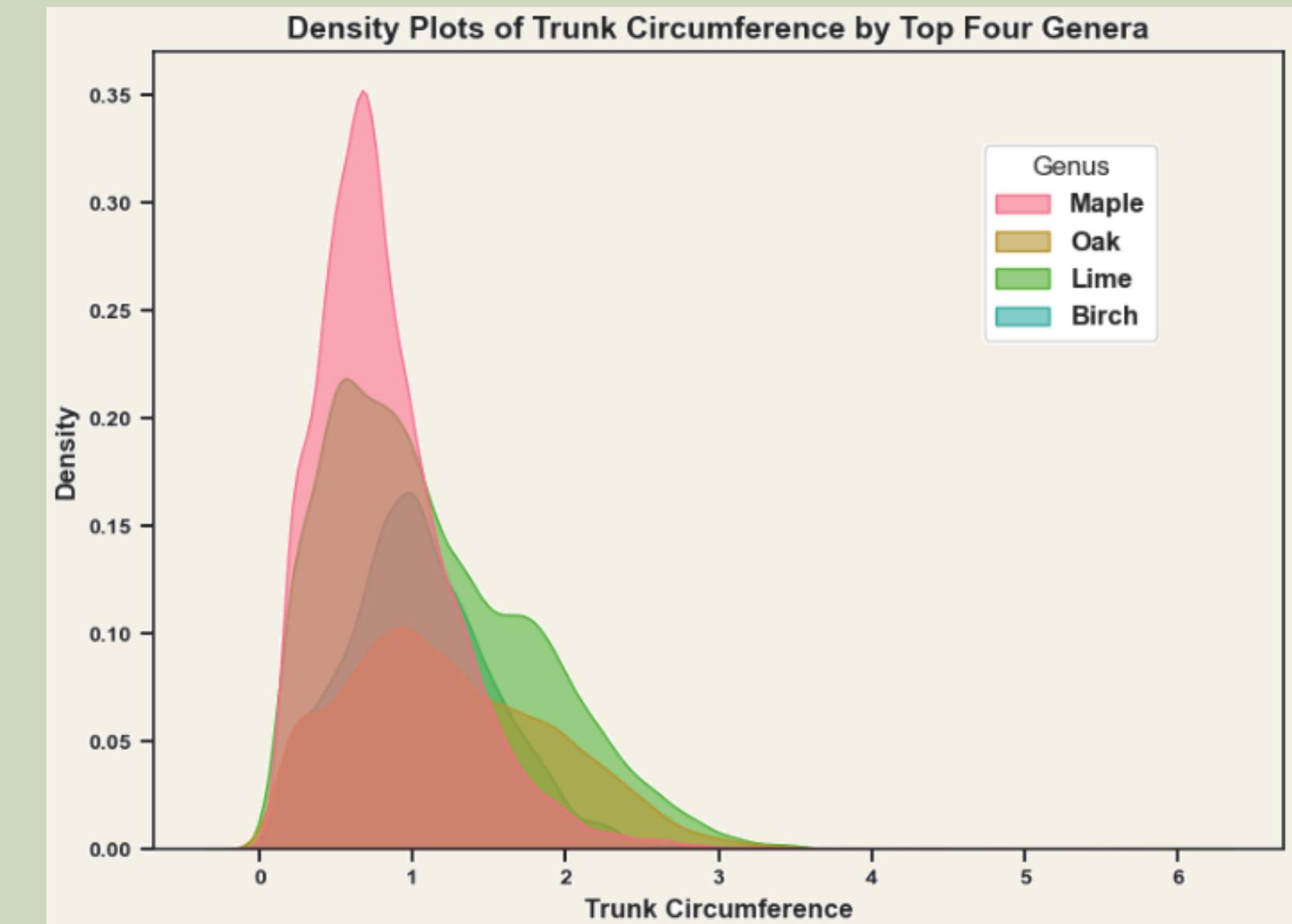
tallest and  
thickest

*Maple trees*

lowest and  
thinnest



**mean**



# EDA - More Cats vs. Measures

- **Sponsored trees** are lower and thinner
- The same is valid for **alley trees**

→ Let's do some **t tests** on this!

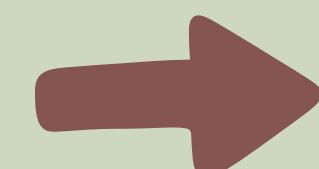




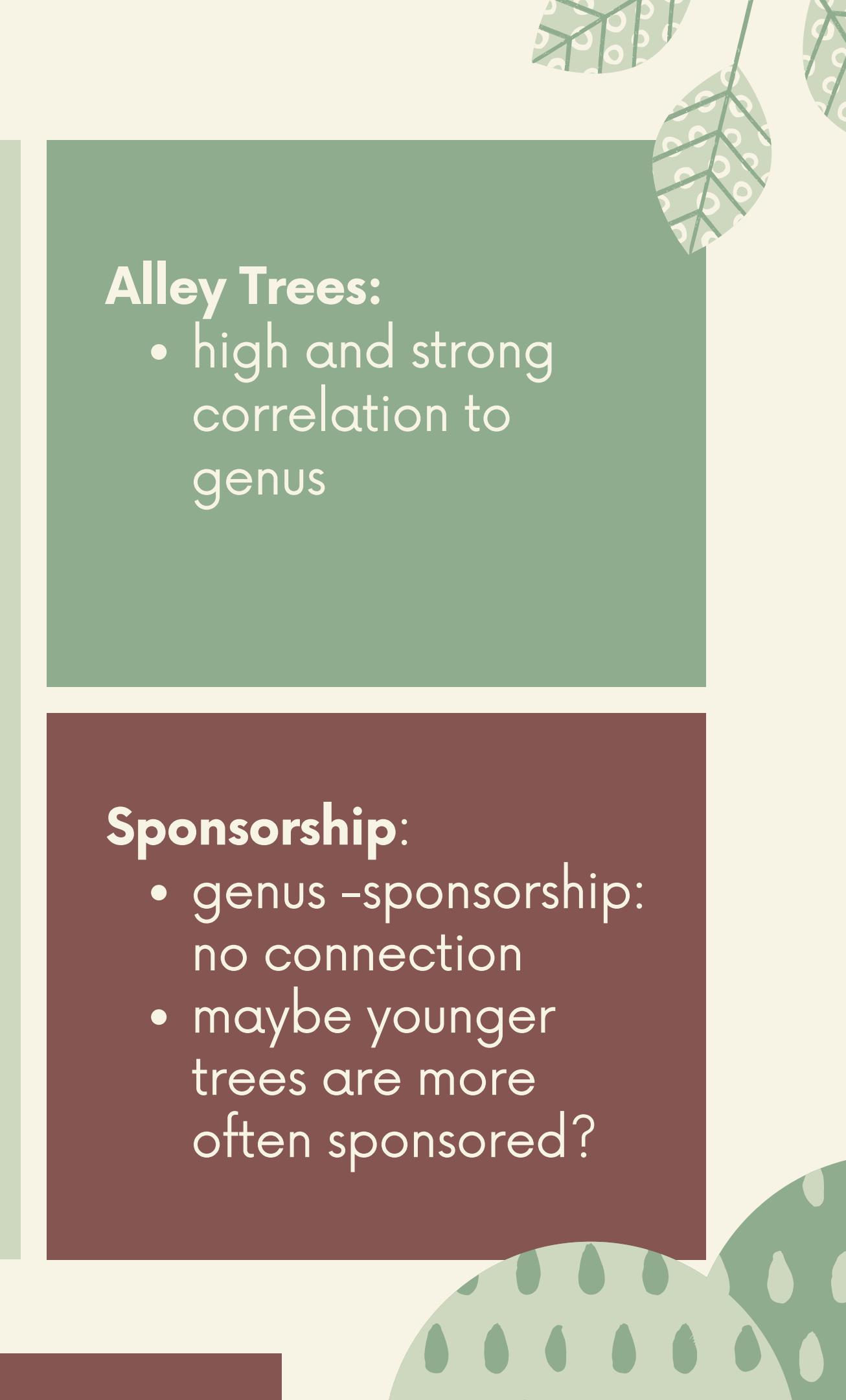
# Inferential Statistics (t tests)

Both relationships are **significant**:

- Sponsored trees are **lower and thinner**, as are alley trees



But **why?**



## Alley Trees:

- high and strong correlation to genus

## Sponsorship:

- genus -sponsorship: no connection
- maybe younger trees are more often sponsored?

# Hypotheses

**H1: Some tree types occur much more often.**

**YES!** Maple and Lime trees especially.



**H2: Some tree types are more often part of an alley.**

**YES!** Lime trees stick out in absolute numbers, Japanese Pagoda, Maidenhair and Hazel in relative numbers.



**H3: Some tree types are more likely to be sponsored.**

**NO!** But a significant relationship could be found regarding tree measurements and sponsorship (age!?).





# Challenges and Obstacles

Making sense of extreme data in tree measurements

Defining sample and pop for t test

Which tests are best for inferential statistics?

Tableau





# **TREES (& me) say**

## **Thank you!**



## **Questions?**

Paula Boks, Midproject, December 2024