Case study: Height

Setup

In [11]:

```
%matplotlib inline
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px

# Custom colors
blue = "#3F83F4"
blue_dark = "#062089"
blue_light = "#8DC0F6"
blue_lighter = "#BBE4FA"
grey = "#9C9C9C"
grey_dark = "#777777"
grey_light = "#B2B2B2"
orange = "#EF8733"
colors_blue = [blue, blue_light]
```

Import data

In [12]:

```
ROOT = "https://raw.githubusercontent.com/kirenz/modern-statistics/main/data/"
DATA = "height.csv"
df = pd.read_csv(ROOT + DATA)
```

In [13]:

```
df.head()
```

Out[13]:

	name	id	height	average_height_parents	gender
0	Stefanie	1	162	161	female
1	Peter	2	163	163	male
2	Stefanie	3	163	163	female
3	Manuela	4	164	165	female
4	Simon	5	164	163	male

In [14]:

```
df.tail()
```

Out[14]:

	name	id	height	average_height_parents	gender
15	Marc	16	166	166	male
16	Ralph	17	166	166	male
17	Tom	18	167	166	male
18	Steven	19	167	167	male
19	Emanuel	20	168	168	male

In [15]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 5 columns):
 #   Column                  Non-Null Count
Dtype
---  ------                  --------------
-----
 0   name                    20 non-null
object
 1   id                      20 non-null
int64
 2   height                  20 non-null
int64
 3   average_height_parents  20 non-null
int64
```

```
 4   gender                  20 non-null
object
dtypes: int64(3), object(2)
memory usage: 928.0+ bytes
```

In [16]:

```
df["name"] = df["name"].astype("category")
df["gender"] = df["gender"].astype("category")
```

Data exploration

In [17]:

```
sns.pairplot(df, hue="gender");
```