

Rules of Data visualization

(asynchronous component of lecture 3)

How to Make a Good Plot

- 1. Show the data.**
- 2. Make patterns easy to see.**
- 3. Display magnitudes honestly.**
- 4. Draw graphics clearly.**

How to Make a Bad Plot

- 1. Hide the data.**
- 2. Make patterns hard to see.**
- 3. Display magnitudes dishonestly.**
- 4. Draw graphics unclearly.**

Mistakes in displaying data

Mistake 1. Hide the data

Mistake 1: Hide the data

How to hide data:

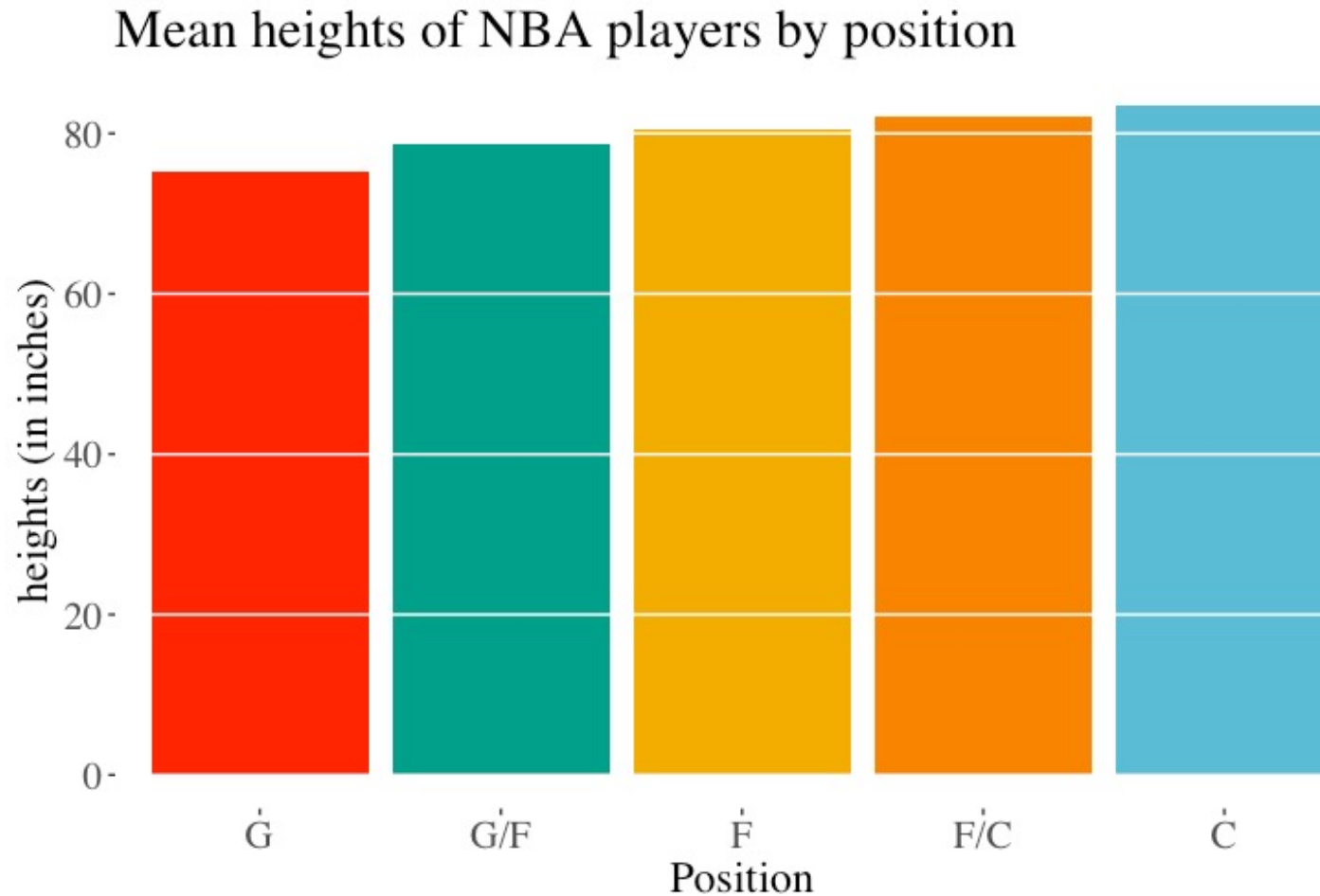
- Provide only statistical summaries.
- Over-plotting.

How to reveal data:

- Present all data points, while allowing all to be seen.

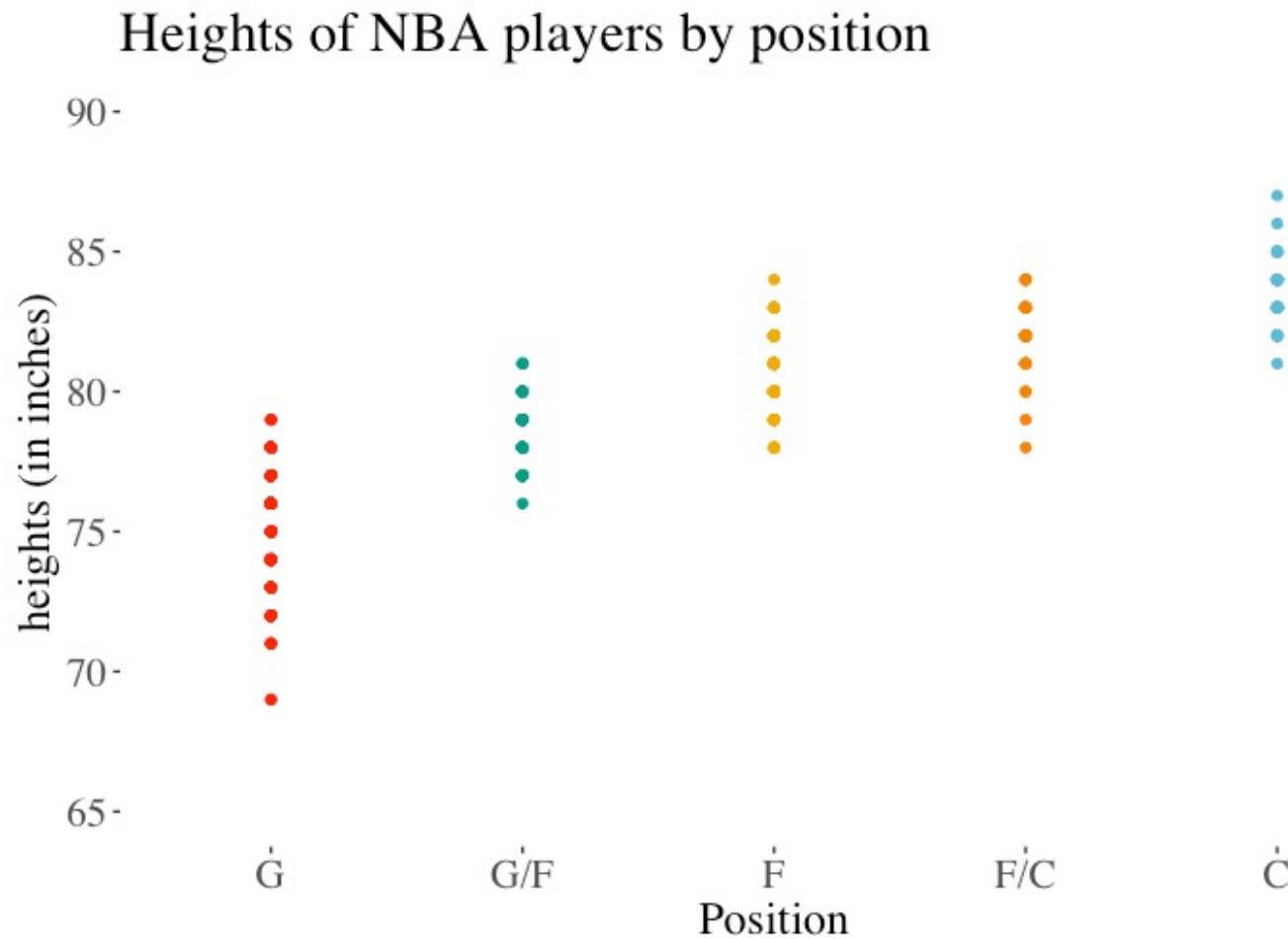
Not Showing Data, Just Summaries

This plot hides the variation within positions.



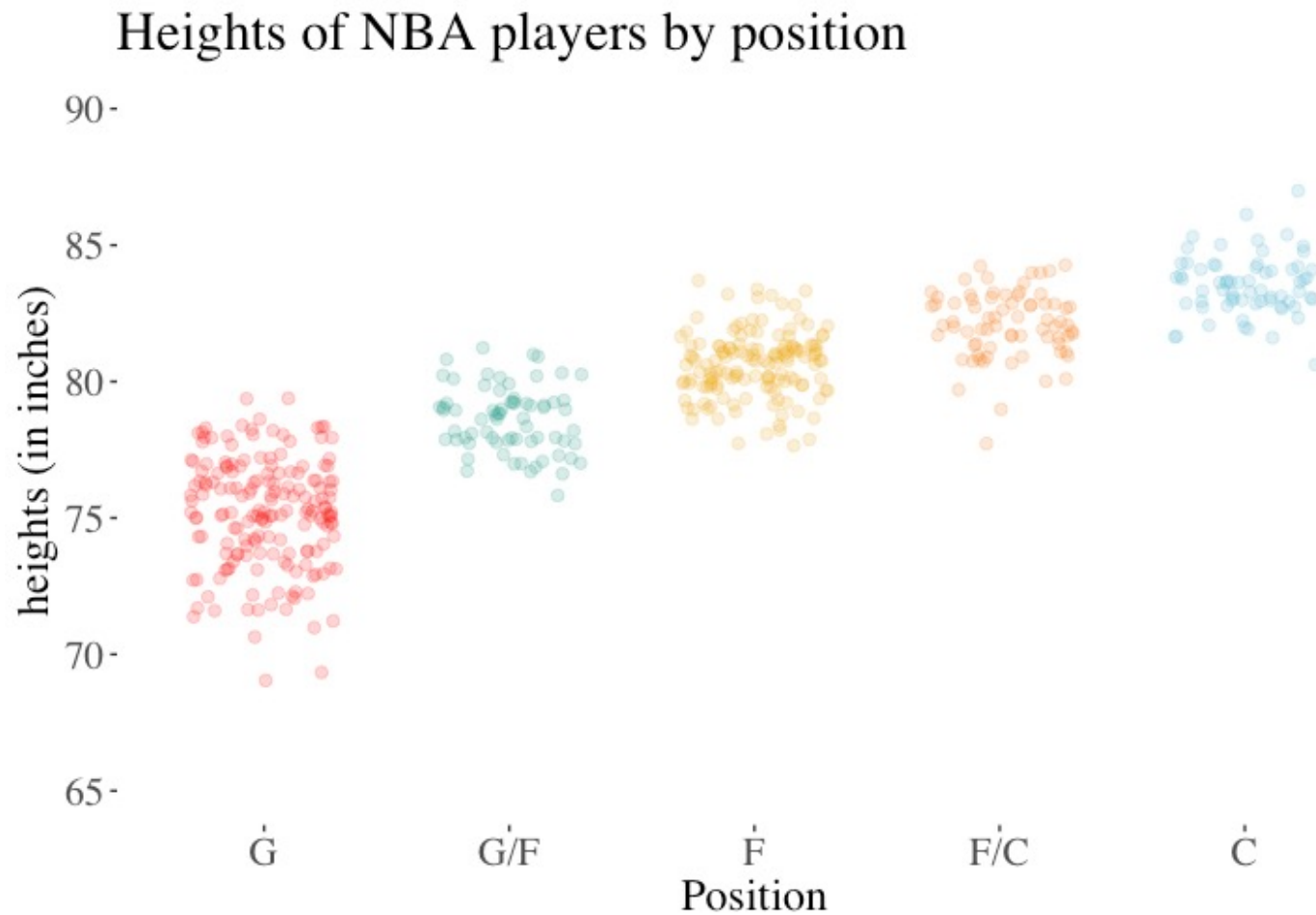
Not Showing Data, Over-Plotting

This plot hides the density of observations.



Showing Data, Jittering

This plot shows all the observations.



Mistakes in displaying data

Mistake 2. Making patterns hard to see

Mistake 2: Making Patterns Hard to See

How to hide patterns:

- Make one plot and call it good.
- Use unreasonable scales.
- Arrange factors nonsensically.

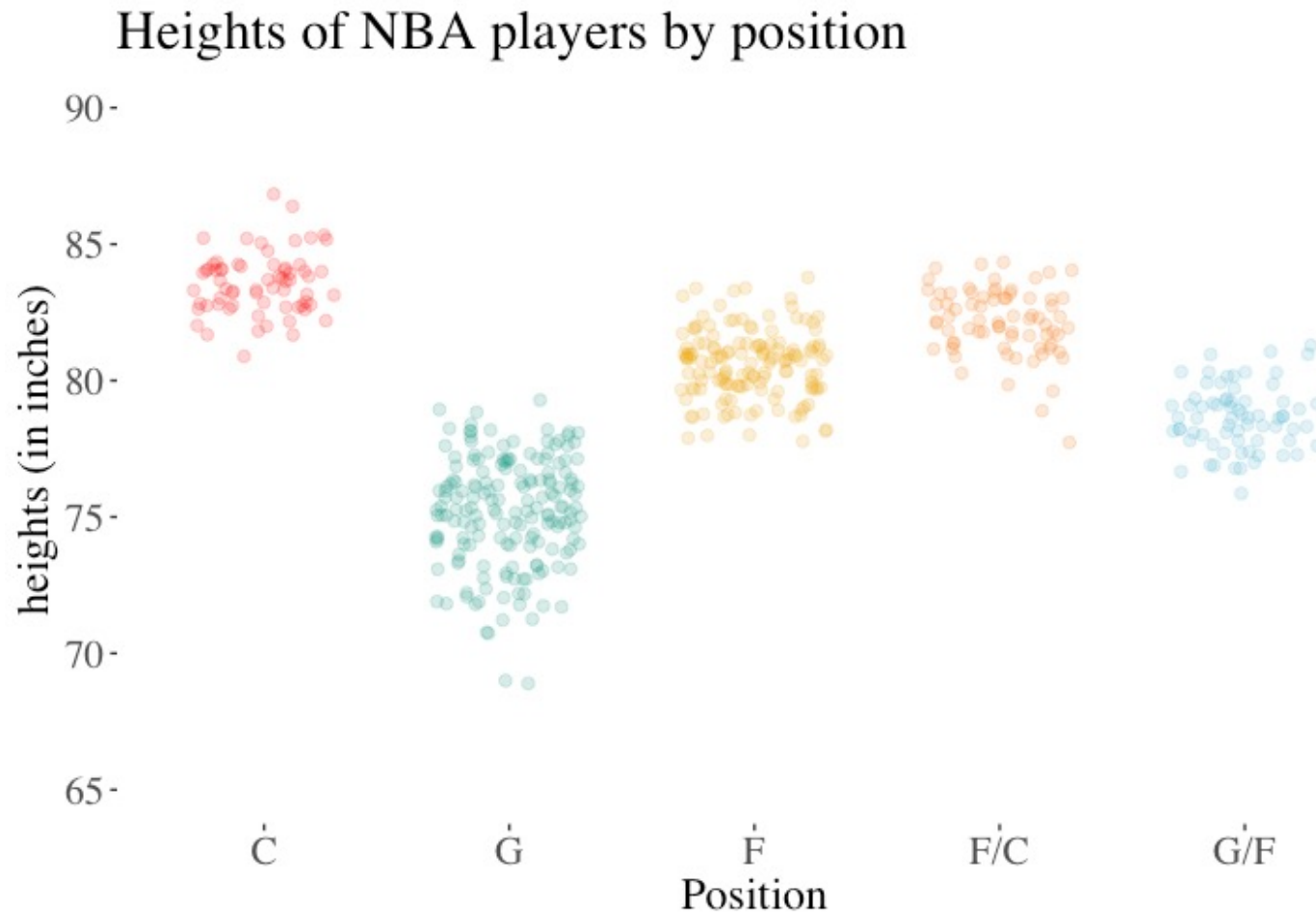
How to reveal patterns:

- Explore multiple potential plots.
- Use appropriate scales.
- Arrange factors meaningfully.

Arrange in order for ordinal, by mean for nominal.

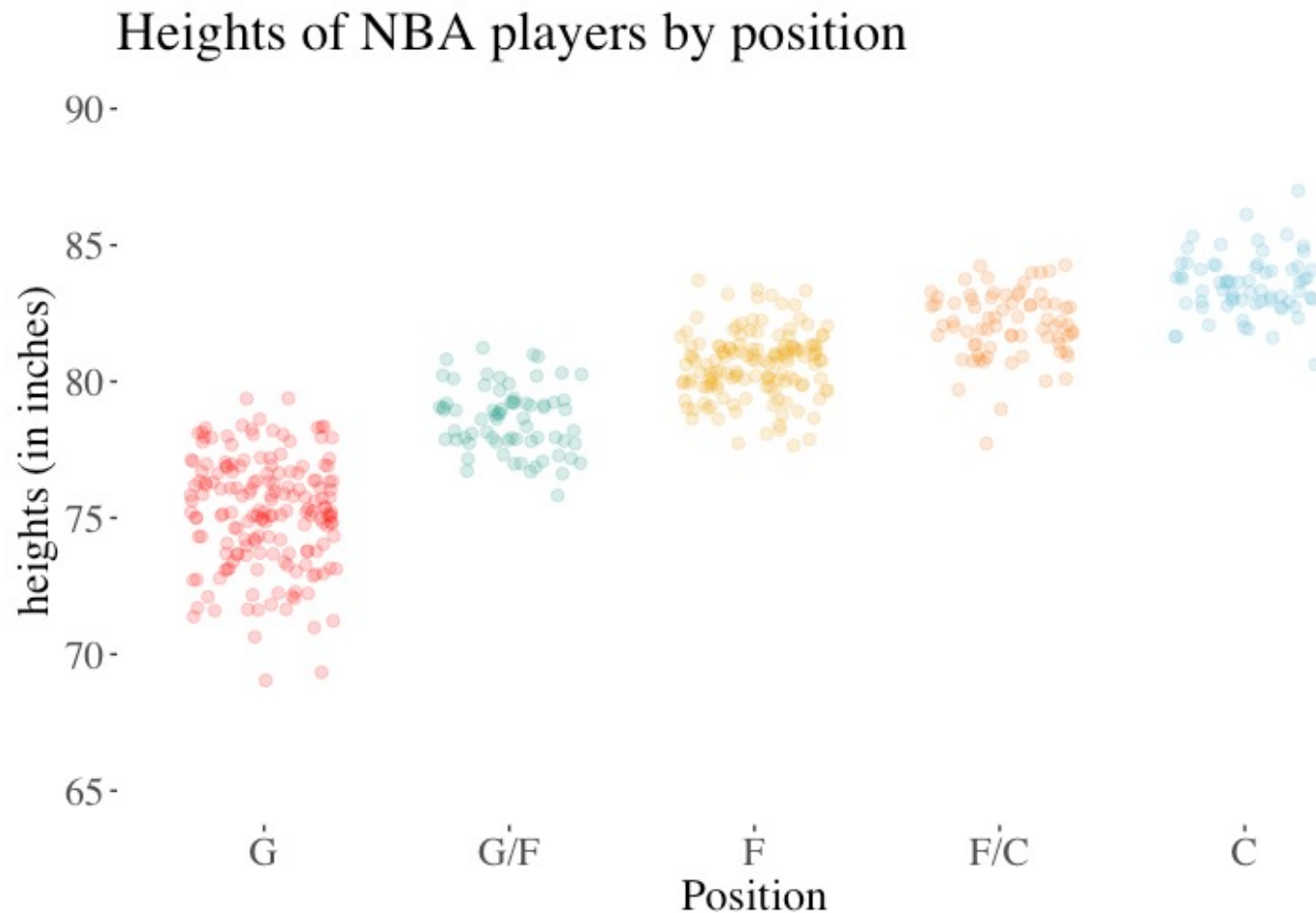
Nonsensical Order Hides Patterns

Non-intuitive ordering of factors hides patterns.



Nonsensical Order Hides Patterns

Intuitive ordering of factors hides patterns.



Bad Axis-Limits Hide Patterns

In this plot, the large scale (limits of the Y-axis) hides the pattern.

