Agency and Uncertainty in Prediction

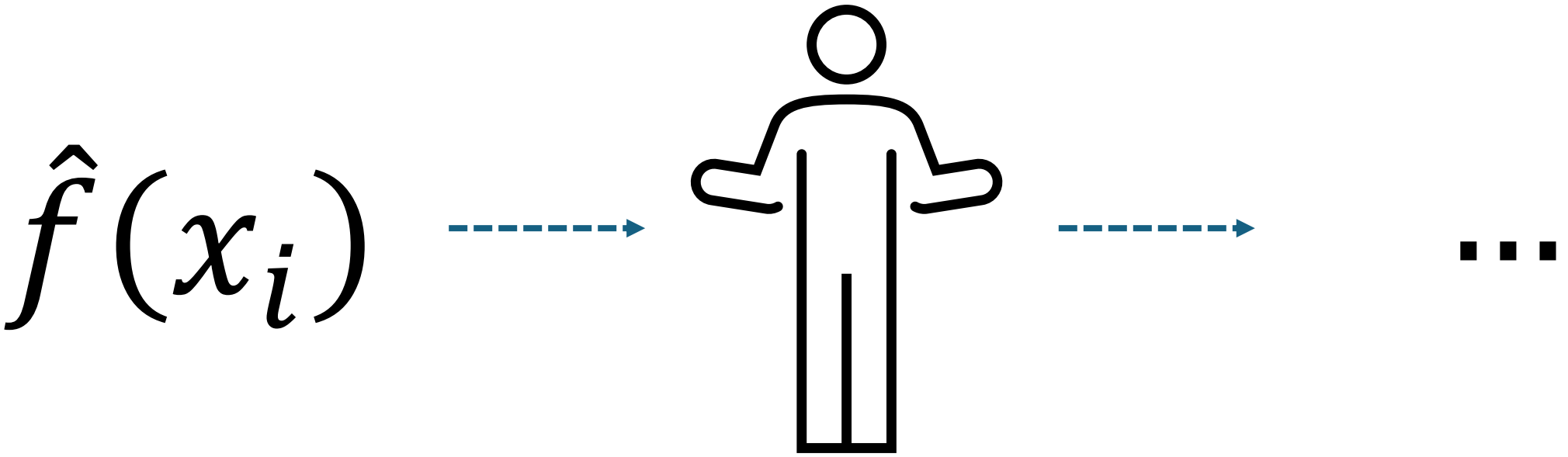
(In regression contexts and with a frequentist orientation)

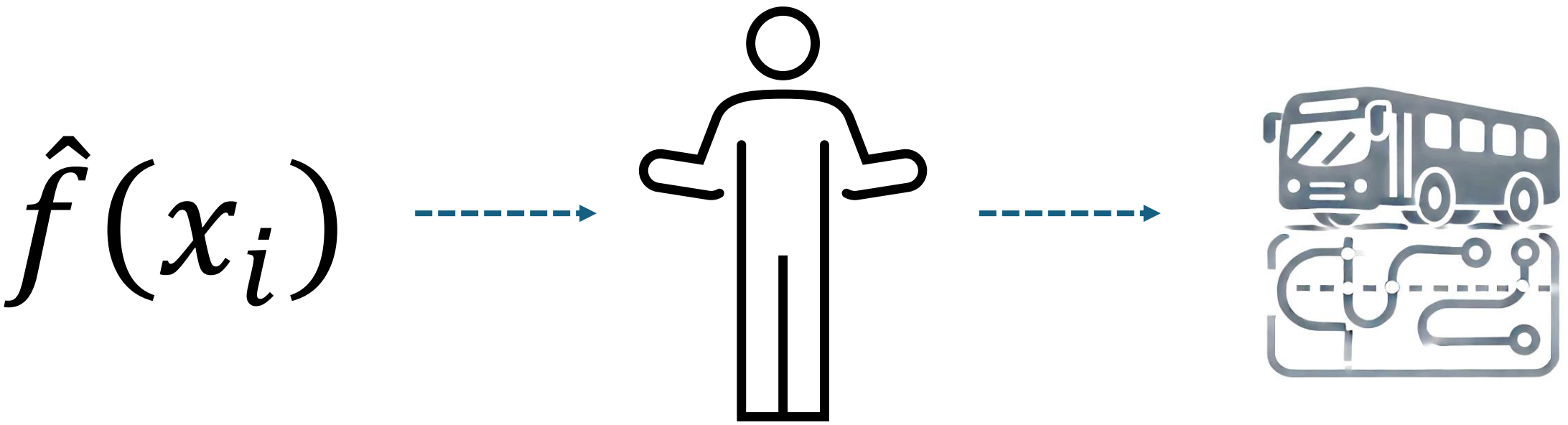
Bryan Shalloway

Data Science @NetApp

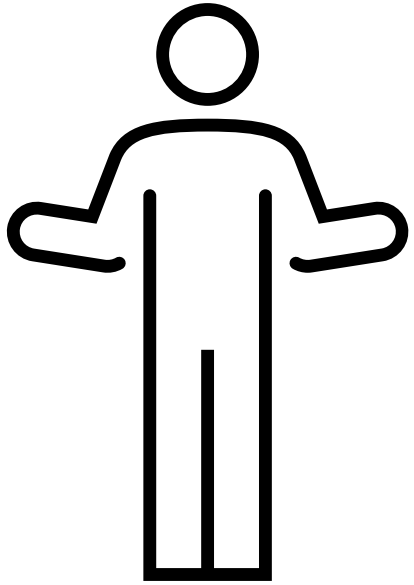
github.com/brshallo/cascadiar-2024


bryanshalloway.com

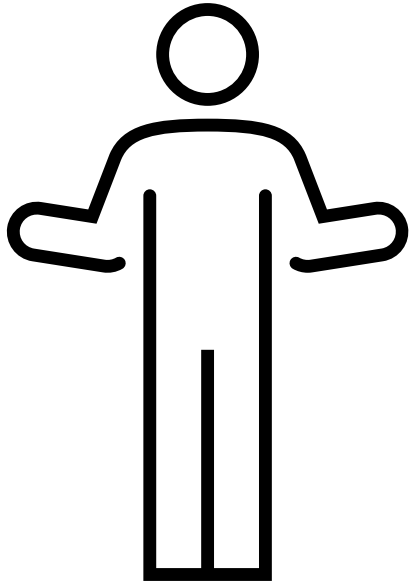

$\hat{f}(x_i)$
...


$\hat{f}(x_i)$

$$\hat{f}(x_i)$$



$$\hat{f}(x_i) \ldots$$



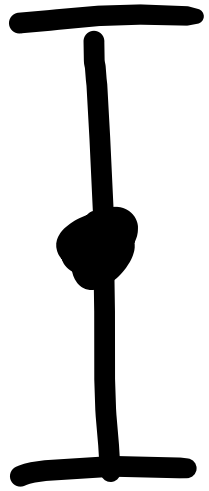
$\hat{f}(x_i)$ ...

- Applicability → Is the model appropriate to use for this observation?
- Uncertainty → What's a reasonable range for this outcome?
- Explainability → What attributes are driving the predicted value?
- ...

Point Estimate

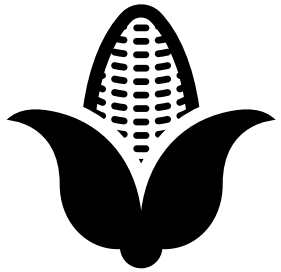
"This car will sell for $12k"

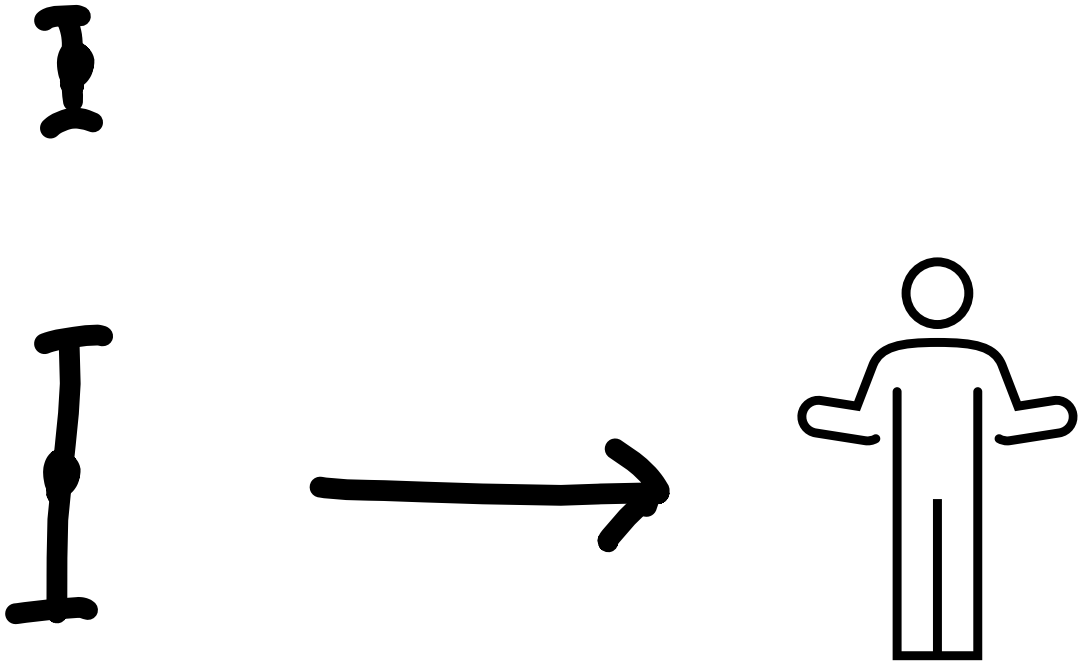
Prediction Interval



"I'm 80% sure this car will sell for between $10k and $14k"

$\hat{f} \longrightarrow \$\$$

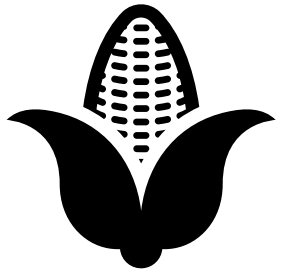


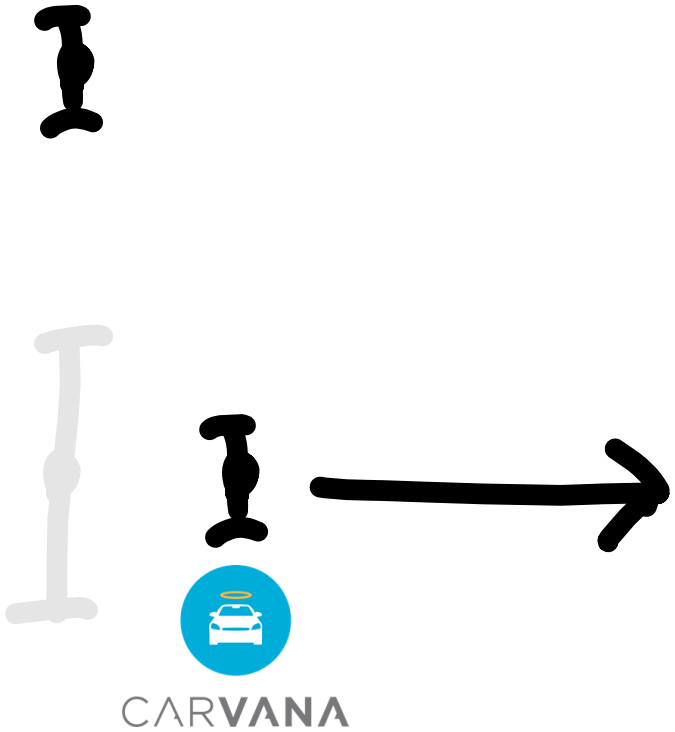


github.com/brshallo/cascadiar-2024

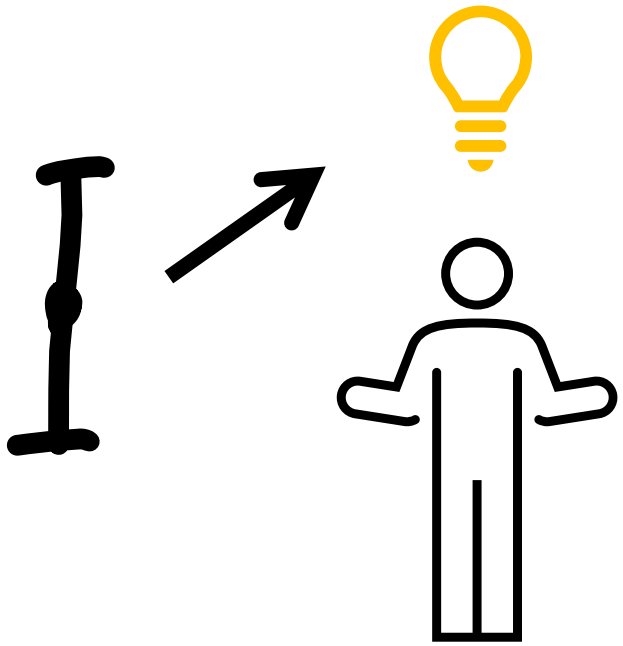
bryanshalloway.com

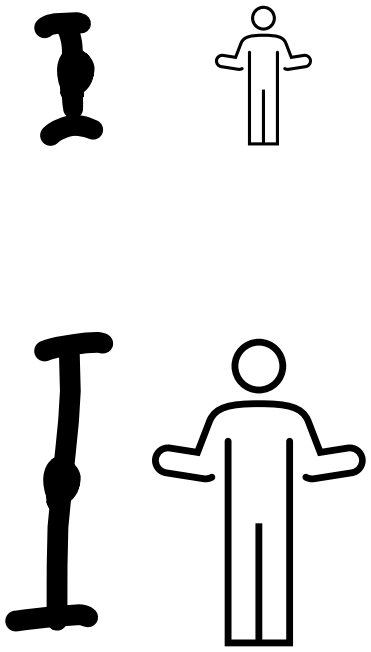
$\hat{f} \longrightarrow \$\$$





Cascadia R Conf
2024





```r
predict(lm_fit, data_test,
        type = "pred_int",
        level = 0.90)
#> # A tibble: 14 × 2
#>    .pred_lower .pred_upper
#>          <dbl>       <dbl>
#>  1        17.6        24.5
#>  2        18.5        25.4
#>  3        18.6        25.5
#>  4        18.8        25.7
#>  5        18.1        25.1
#>  6         4.19       11.2
#>  7         2.72        9.67
#>  8        17.0        23.9
#>  9        18.4        25.3
#> 10        18.4        25.3
#> 11        18.4        25.3
#> 12        17.7        24.6
#> 13         5.27       12.2
#> 14         4.26       11.2
```

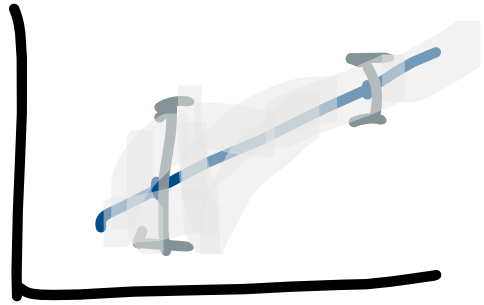
```
predict(lm_fit, data_test,
        type = "pred_int",
        level = 0.90)
#> # A tibble: 14 × 2
#>    .pred_lower .pred_upper
#>          <dbl>       <dbl>
#>  1        17.6        24.5
#>  2        18.5        25.4
#>  3        18.6        25.5
#>  4        18.8        25.7
#>  5        18.1        25.1
#>  6         4.19       11.2
#>  7         2.72        9.67
#>  8        17.0        23.9
#>  9        18.4        25.3
#> 10        18.4        25.3
#> 11        18.4        25.3
#> 12        17.7        24.6
#> 13         5.27       12.2
#> 14         4.26       11.2
```

...weaknesses

...weaknesses

- No guarantee of coverage

- Other model types



- Assumptions

...

desired...

- Coverage guaranteed
- Model Agnostic
- Assumption free (and flexible)

...weaknesses

- No guarantee of coverage
- Other model types
- Assumptions
- ...

Quantile regression + adjusted (based on holdout)

(AKA Conformalized Quantile Regression)

Go here next:

- `probably::int_conformal_quantile()`
  Conformal Inference with Tidymodels - posit::conf(2023); *Kuhn* (https://youtu.be/vJ4BYJSg734?si=cjpXabfmAad1FuBK)
- A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification; *Angelopoulos, Bates* (https://people.eecs.berkeley.edu/~angelopoulos/blog/posts/gentle-intro/)
- Introduction To Conformal Prediction With Python; *Molnar*