ADVANCED DATA VISUALIZATION AND REPORTING

SETTING THE STAGE

data-action-lab.com

OUTLINE

1. Data Visualization with ggplot2

2. An Introduction to Dashboards

R GRAPHICS

As of 2018, there are 4 graphical systems available in R:

- *base*

- *grid*

- *lattice*

- *ggplot2*

Access to the 4 systems differ: *base*, *grid*, *lattice* are included in the base installation; *grid*, *lattice*, *ggplot2* have to be loaded explicitly before being used.

A GGPLOT2 PRIMER

ggplot2 is a set of tools that map data to visual display elements, and that allow the user to control the fine details of plot display.

Most important aspect: *ggplot2* can be used to think about the **logical structure** of the plot.

A *ggplot2* graph has 2 main components (and optional terms):

- aesthetic mappings (**aes –** connections between data and plot elems.)

- plot geometry (**geom –** specifies the type of plot)

- *facets, *coordinates, *scales, *labels, *guides, etc.

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

GGPLOT2 GRAMMAR

1. Tidy Data

`p ← ggplot(data = gapminder, ...`

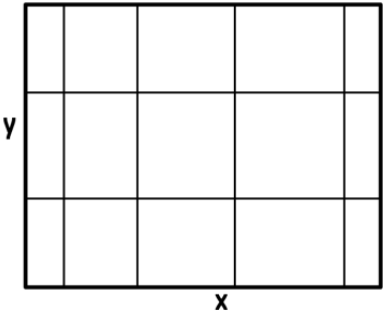| gdp	lifexp	pop	continent
340	65	31	Euro
227	51	200	Amer
909	81	80	Euro
126	40	20	Asia

2. Mapping

`p ← ggplot(data = gapminder, mapping = aes(x = gdp, y = lifexp, size = pop, color = continent))`

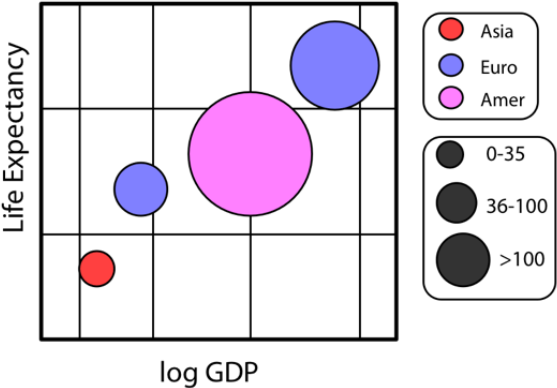### 3. Geom

`p + geom_point()`

4. Co-Ordinates & Scales

`p + coord_cartesian() + scale_x_log10()`

5. Labels & Guides

`p + labs(x = "log GDP", y = "Life Expectancy", title = "A Gapminder Plot")`

A Gapminder Plot

GGPLOT2 GRAMMAR – GEOMS

The data source and variables to be plotted are specified *via* `ggplot()`.

The various geom functions specify **how** these variables are to be visually represented

- using points, bars, lines, shaded regions, etc.

There are currently 37 available geoms.

GGPLOT2 GRAMMAR – GEOM()

```r
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
  geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
  geom_boxplot()
```
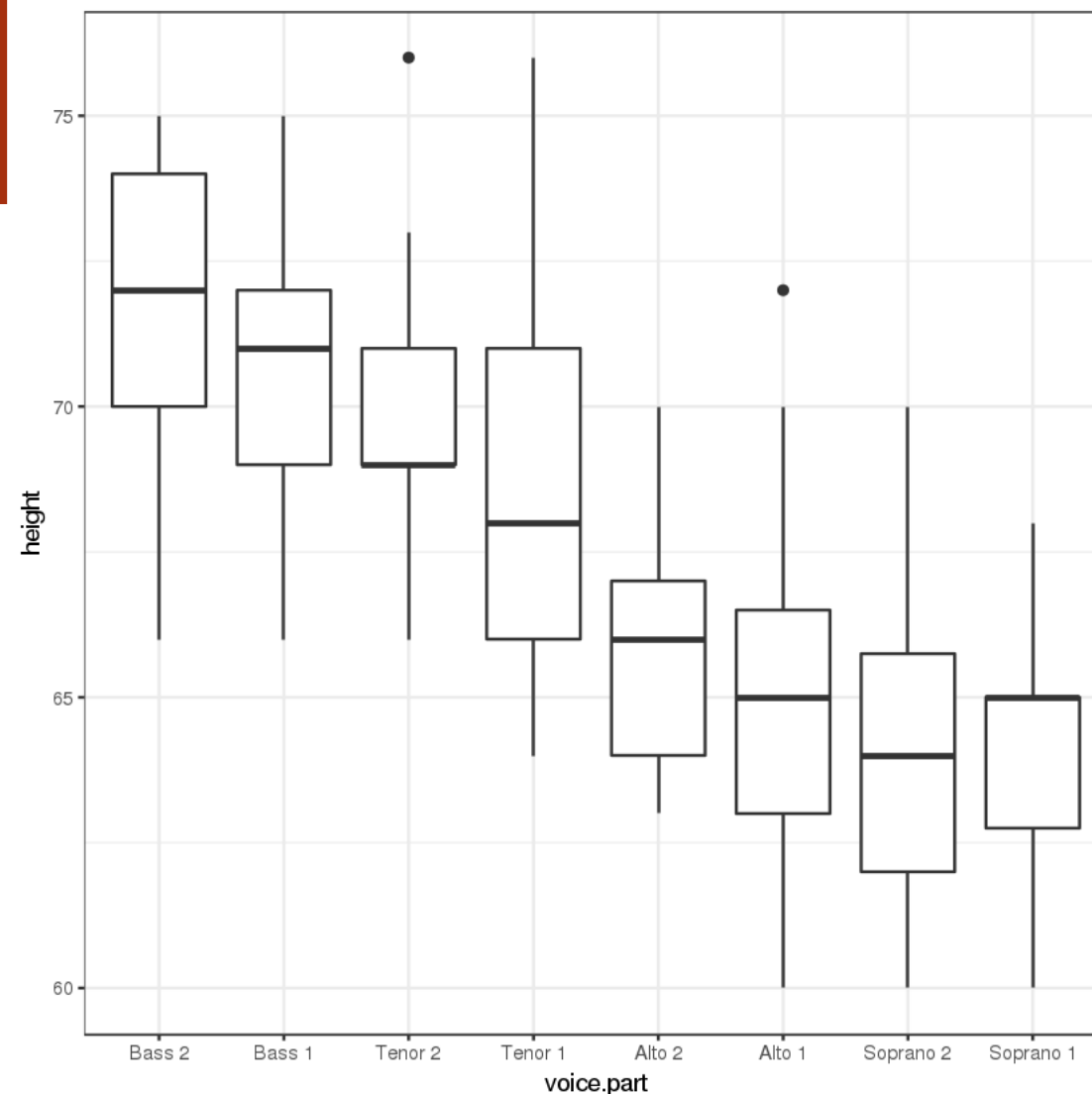
GGPLOT2 GRAMMAR – GEOM()

```r
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
    geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
    geom_boxplot()
```
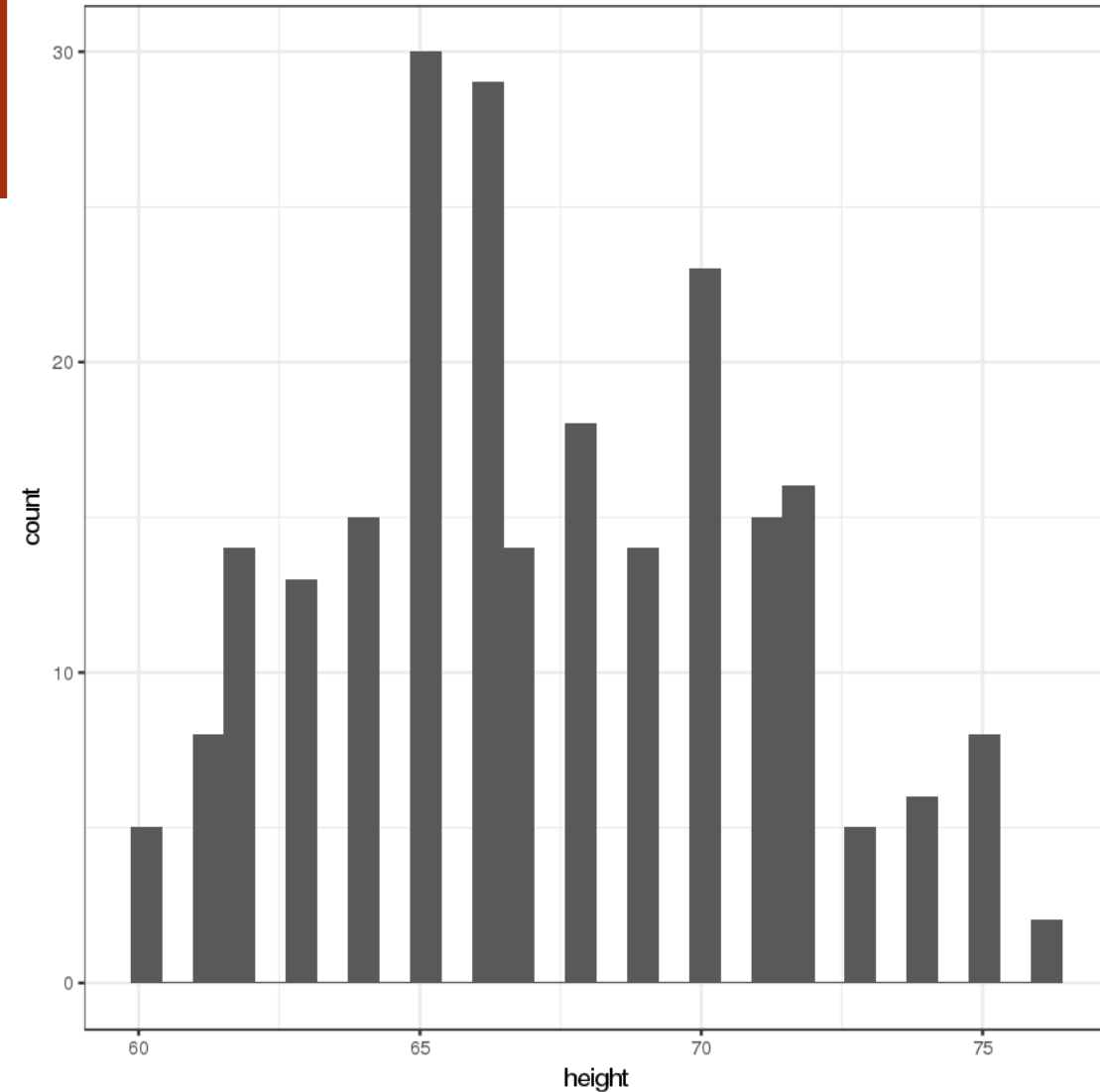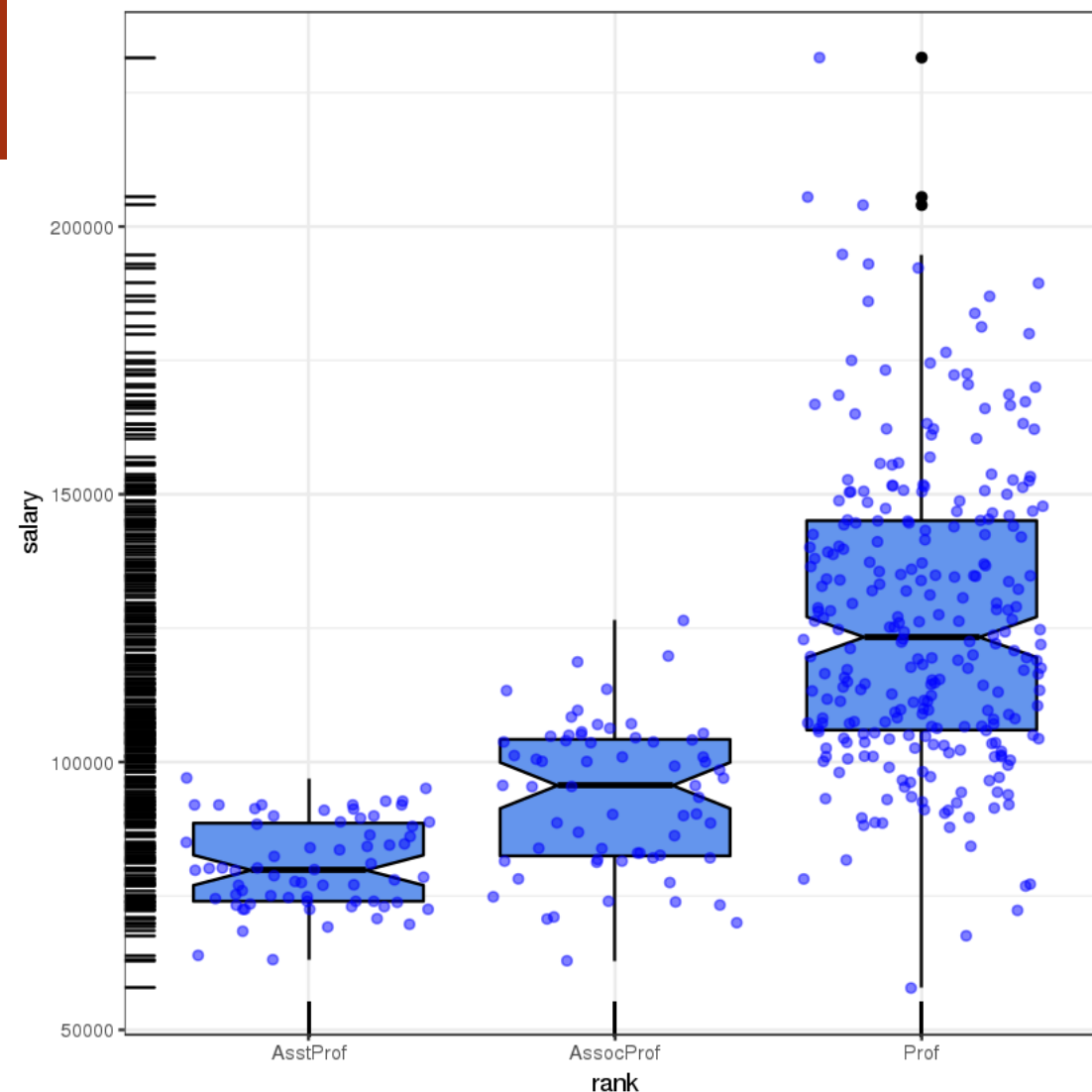
GGPLOT2 GRAMMAR – GEOM()

```r
library(ggplot2)
data(Salaries, package="car")
# Using data on salaries of a sample of
# US university professors (2018-2019)
# var: rank, sex, yrs.since.phd, yrs.service, salary

ggplot(Salaries, aes(x=rank, y=salary)) +
    geom_boxplot(fill="cornflowerblue",color="black", notch=TRUE) +
    geom_point(position="jitter", color="blue", alpha=.5) +
    geom_rug(side="l", color="black")
```

Aesthetics refer to the displayed attributes of the data.

They map the data to an attribute (such as the size or shape of a marker) and generate an appropriate legend.

Aesthetics are specified with the `aes()` function.

Aesthetics can be specified within the data function or within a geom. If they're specified within the data function then they apply to all specified geoms.

GGPLOT2 GRAMMAR – AESTHETICS

The aesthetics **available to** `geom_point()` **(scatterplot), as an example, are:**

- `x, y, alpha, color, fill, shape, size`

Important difference between specifying characteristics (like colour and shape) inside and outside the `aes()` function
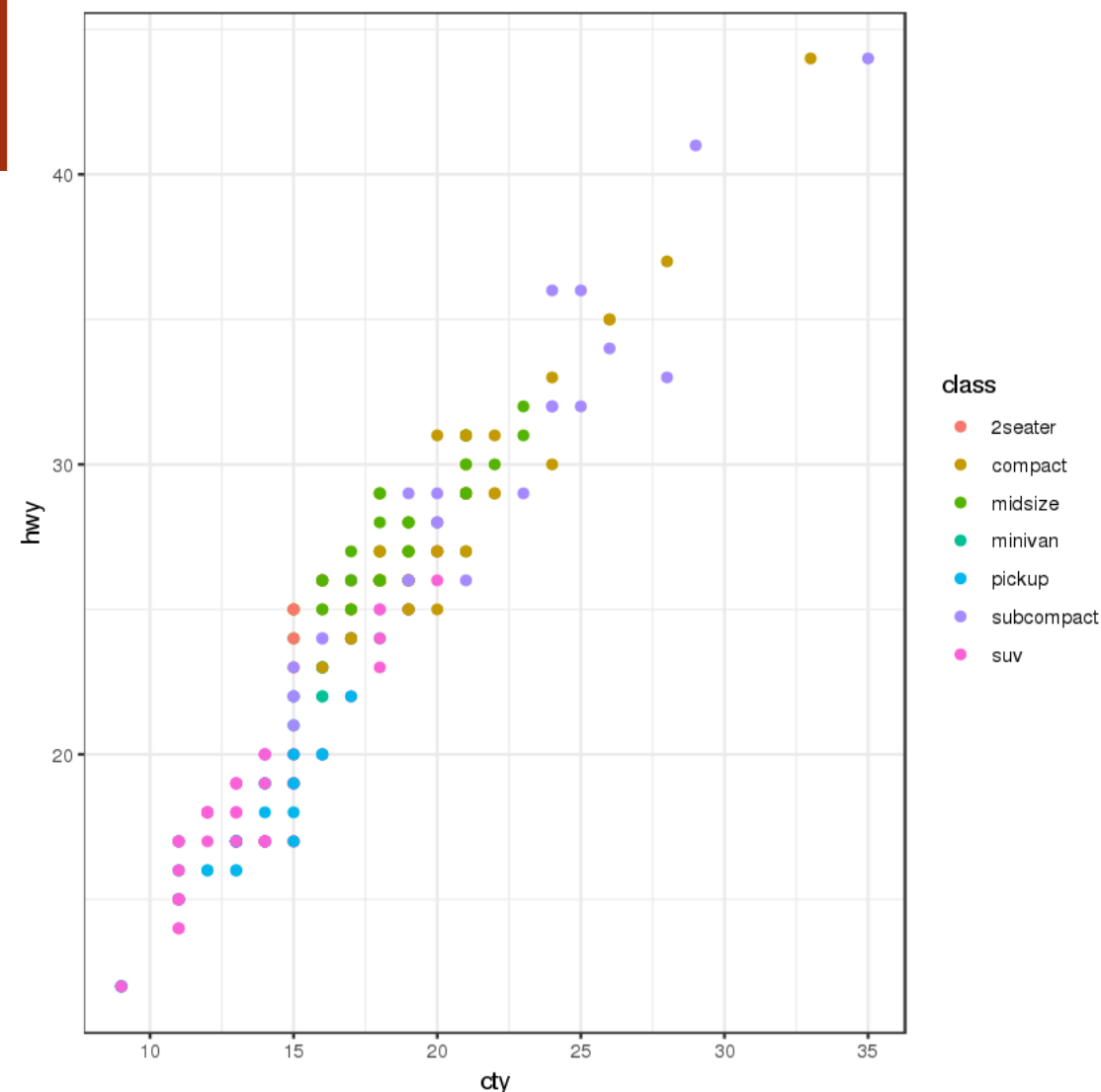
- inside: assigned colour or shape automatically based on the data.

- outside: not mapped to data.

GGPLOT2 GRAMMAR – AES()

```r
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(colour = "red")
```

GGPLOT2 GRAMMAR – AES()

```r
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
   geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
   geom_point(colour = "red")
```
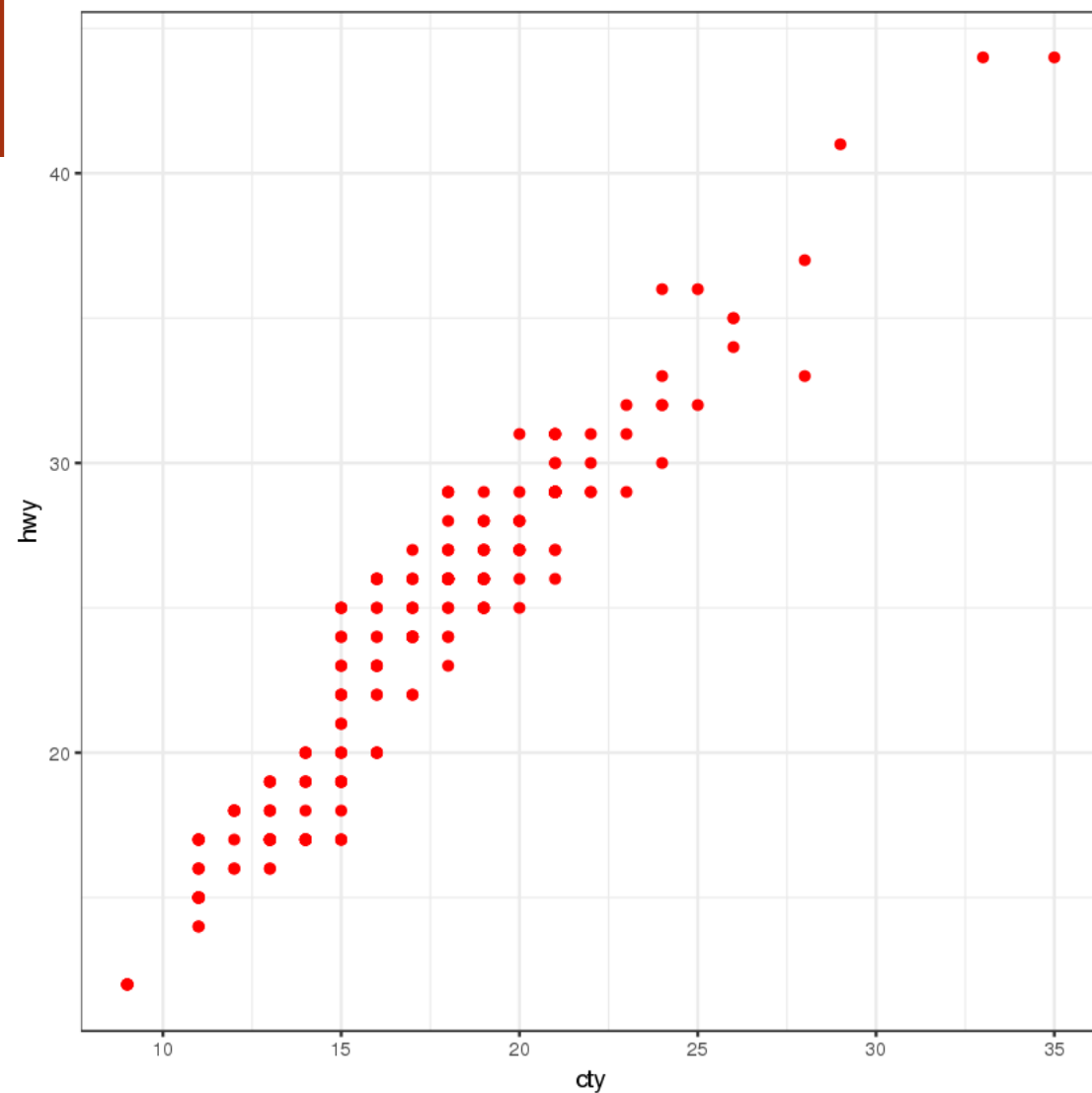
GGPLOT2 GRAMMAR – FACETS

In *ggplot2* parlance, small multiples are referred to as **facets**:

- `facet_wrap(), facet_grid()`

By default, all panels (one for each factor) share the same axes (scale-wise).

Separating the graph into a sequence of smaller, side-by-side plots makes it easier to enact comparisons.

GGPLOT2 GRAMMAR – FACET_WRAP()

```r
data(singer, package="lattice")
library(ggplot2)
ggplot(data=singer, aes(x=height)) +
    geom_histogram() +
    facet_wrap(~voice.part, nrow=4)
```
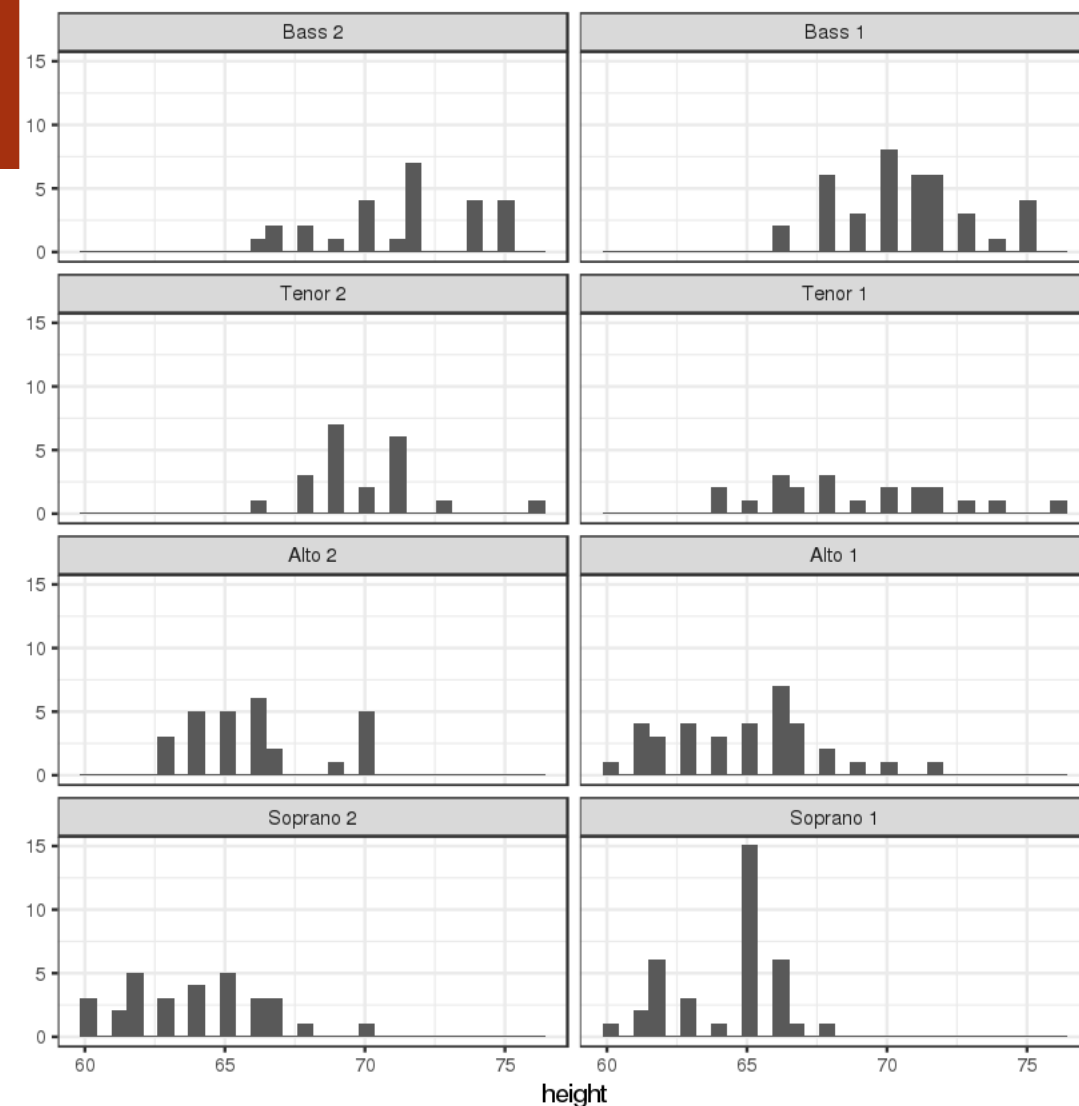
GGPLOT2 GRAMMAR – FACET_GRID()

```
data(Salaries, package="car")
library(ggplot2)
ggplot(Salaries, aes(x=yrs.since.phd,
 y=salary, color=rank, shape=rank)) +
    geom_point() +
    facet_grid(~sex)
```
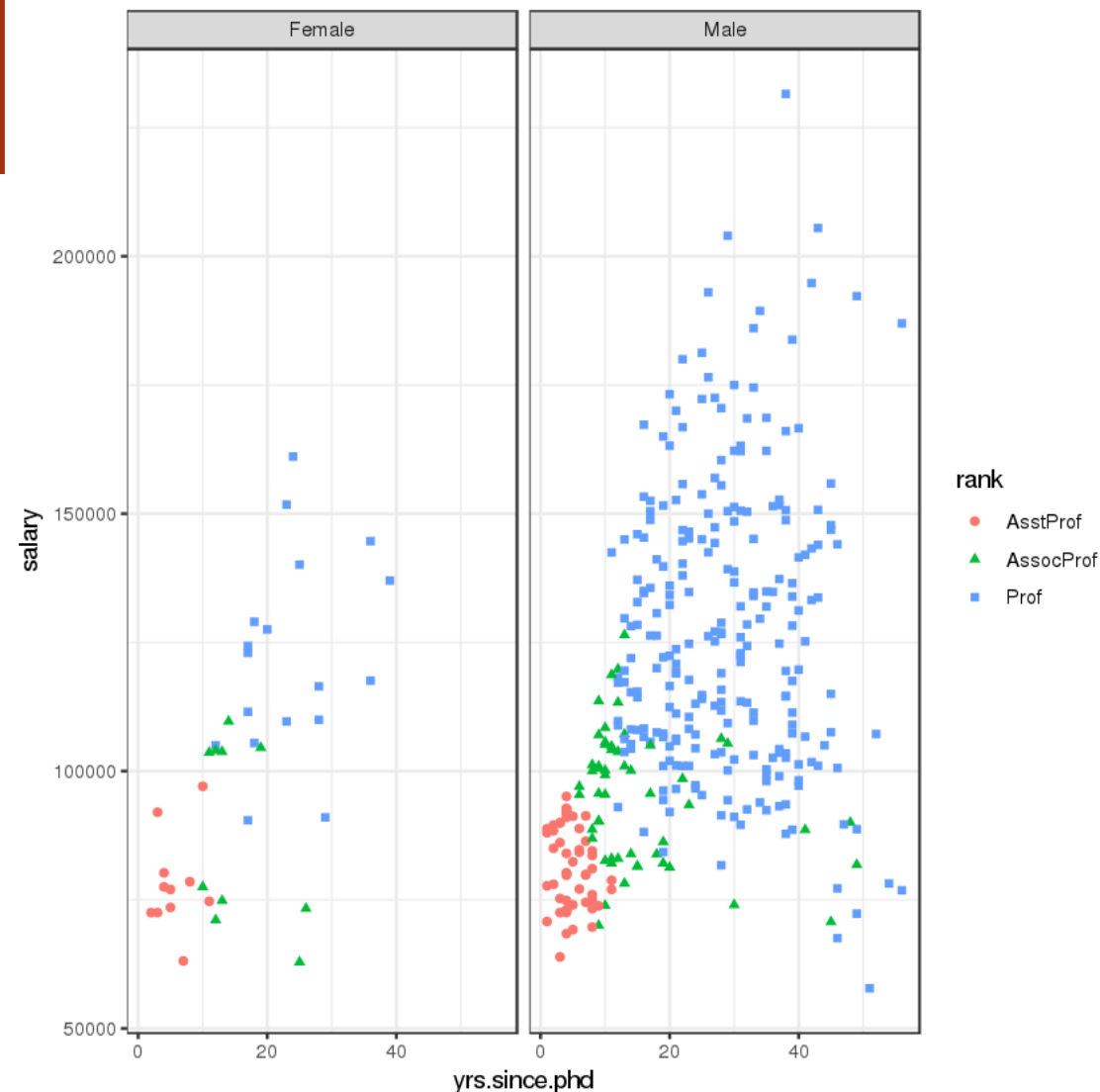
REPORTING AND DEPLOYMENT

An analysis can only be as good as how it is **communicated** and/or **deployed**.

Crucial Questions:

- Who is in receipt of the report(s)?

- How are the workflows deployed into production?

- Can data insights be turned into useful policies?

Automatic reporting should be audited and validated **regularly**.

REPORTING AND DEPLOYMENT

Communication should occur at various stages of the project, not solely upon completion:

- keep sponsors / clients aware of broad lines

- technical details may be avoided, but documented nonetheless

Ideal scenario: analysis software is also reporting software

- minimizes human error related to cut-and-paste

- removes the need for keeping analysis and reporting separate

- makes sharing the work with other project member easier

Simplify the process further by deploying directly to the Web.

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

DASHBOARDS

A **dashboard** is any visual display of data used to monitor conditions and/or facilitate understanding.

Examples:

- interactive display that allows people to explore motor insurance claims by city, province, driver age, etc.

- PDF showing key audit metrics that gets e-mailed to a Department's DG on a weekly basis.

- wall-mounted screen that shows call centre statistics in real-time.

- mobile app that allow hospital administrators to review wait times on an hourly- and daily-basis for the current year and the previous year.

SOME QUESTIONS TO CONSIDER

In a car's dashboard, a small number of **key indicators** (speed, gasoline level, lights, etc.) need to be understood **at a glance**. A dashboard design that does not take these two characteristics under consideration can have catastrophic consequences.

The following questions need to be answered prior to the dashboard being designed:

- Who is the dashboard's **consumer**?

- What **story** does the dashboard tell?

- What data (categories) will be used?

- What will **appear** on the dashboard?

- How can the dashboard **help** the consumer?

IDLEWYLD   Sysabee   DAVHILL

data-action-lab.com
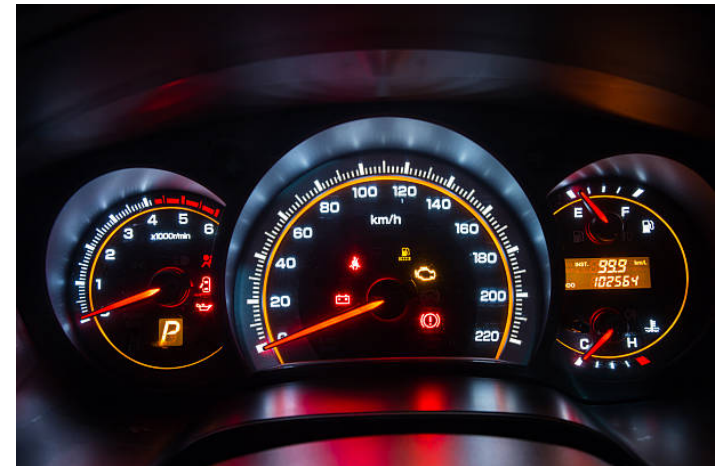
DASHBOARD DESIGN GUIDELINES

Nick Smith suggests the following 6 Golden Rules:

- **Consider the audience** (who are you trying to inform? does the DG really need to know that the servers are operating at 88% capacity?)

- **Select the right type of dashboard** (operational, strategic/executive, analytical)

- **Group data logically, use space wisely** (split functional areas: product, sales/marketing, finance, people, etc.)

- **Make the data relevant to the audience** (scope and reach of data, different dashboards for different departments, etc.)

- **Avoid cluttering the dashboard** (present the most important metrics only)

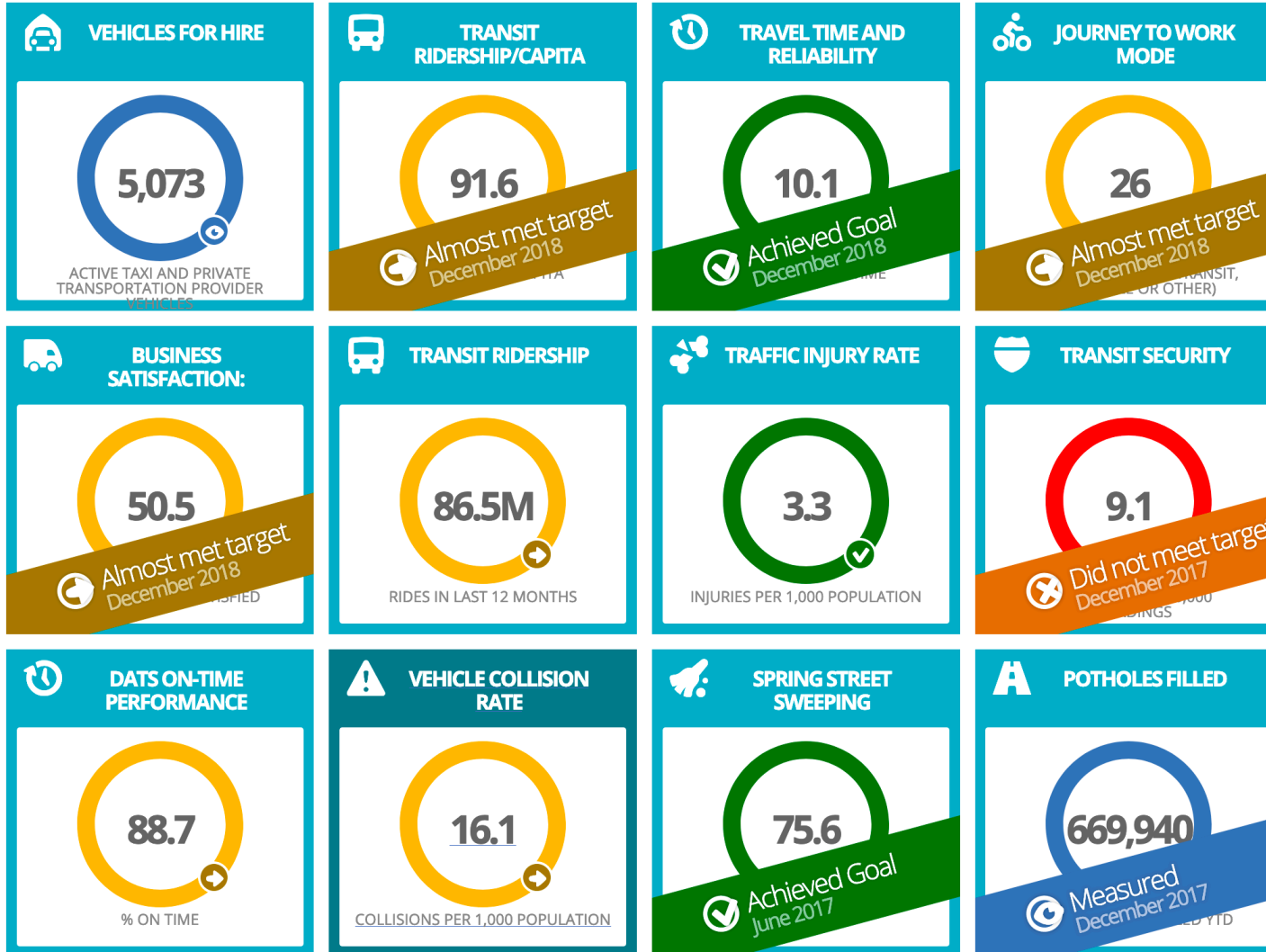- **Refresh your data at the right frequency** (real-time, daily, weekly, monthly, etc.)

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

Course Metrics

Students

52

S S F S F S US F S F F
'12 '13 '14 '15 '16
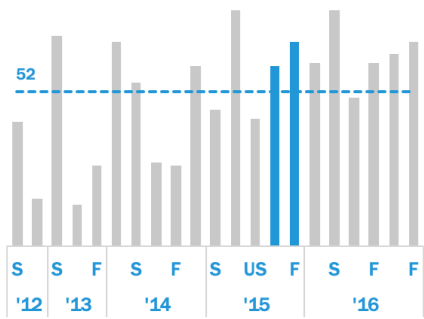
1097
Total Students in five years

Enrollments

388
299
240
112
58

'12 '13 '14 '15 '16
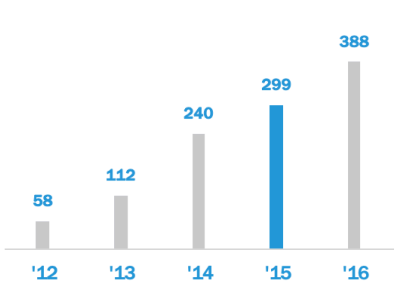
687
Total Students in 2015-2016

Classes

6
5 5
3
2

'12 '13 '14 '15 '16
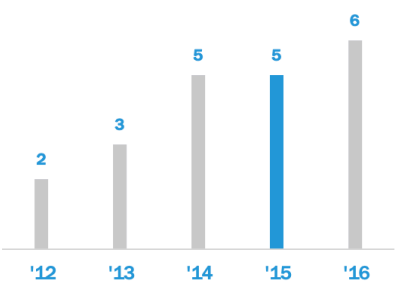
21
Total Classes in five years

Ratings

8

4

0

S F S US F S S US F F S S US F F
'12 '13 '14 '15
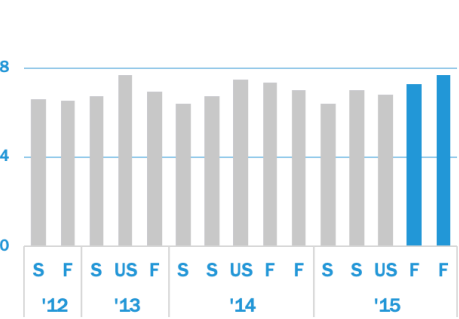
7.7 of **8**
Most recent instructor rating (out of 8.0)

Semesters	Questions	●BANA ❘College ●Shaffer	Ratings

2015 Fall Semester 001

I developed specific skills and competencies	6.9
Overall, this was an excellent course	7.1
The instructor communicated clearly	7.4
The Instructor graded fairly	7.5
The instructor was well organized	7.3
The instructor interacted well with students	7.3
Overall, this instructor was excellent	7.3

2015 Fall Semester 002

I developed specific skills and competencies	7.2
Overall, this was an excellent course	7.4
The instructor communicated clearly	7.6
The Instructor graded fairly	7.6
The instructor was well organized	7.5
The instructor interacted well with students	7.7
Overall, this instructor was excellent	7.7

2 3 4 5 6 7 **Out of 8**

Course Metrics Dashboard created by Jeffrey A. Shaffer. Data from University of Cincinnati Course Evaluations. Blue indicates the 2 most recent rating periods.
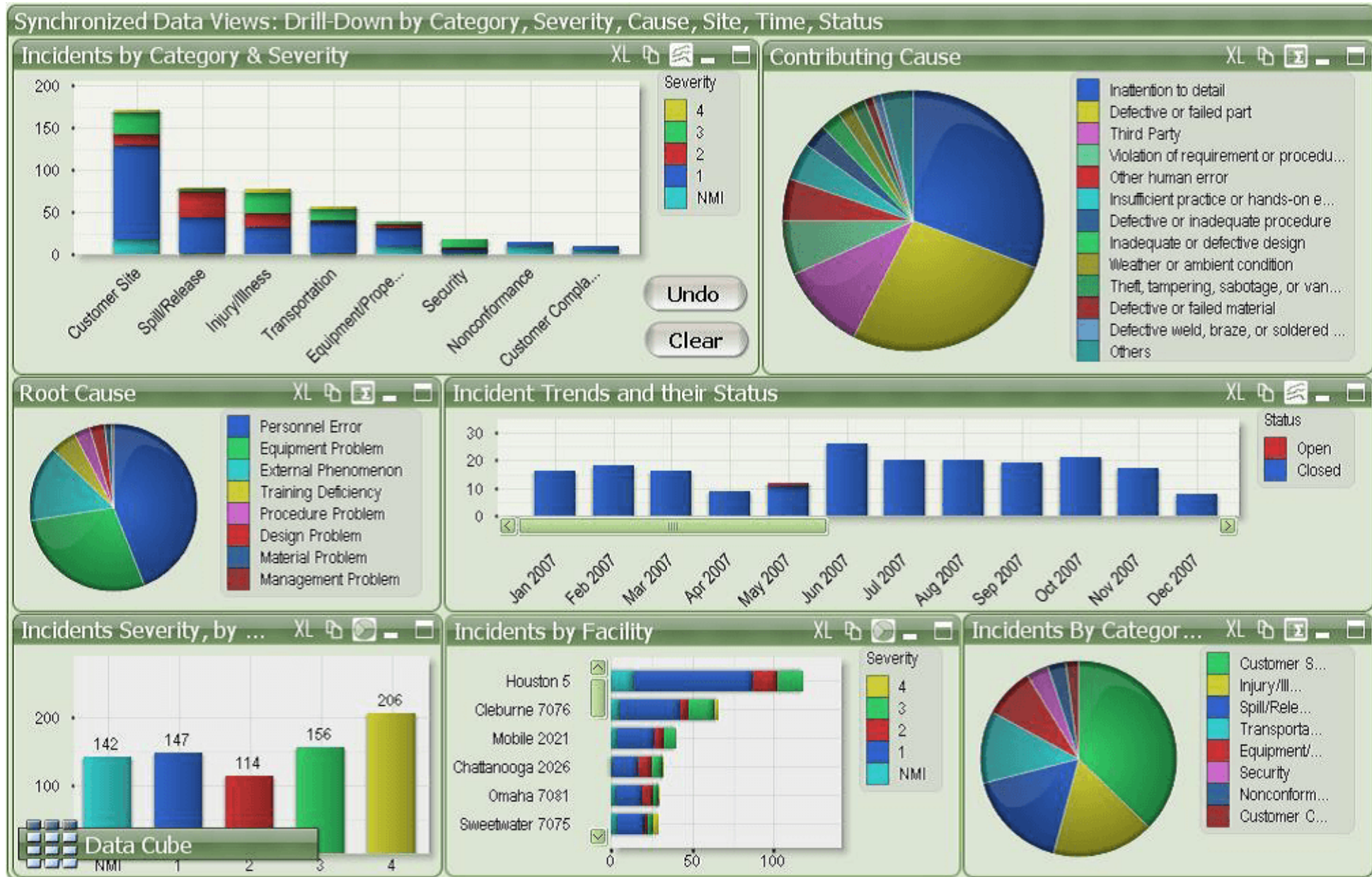
COURSE METRICS DASHBOARD – STRENGTHS

Easy-to-see key metrics

Simple color scheme

Potential to be static or interactive

Both overview and details are clear

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

What-If Analysis: Impact of Minimum Wage [https://bigbookofdashboards.com/dashboards.html]

Developed by Matt Chambers
http://sirvizalot.blogspot.com/

Proposed Minimum Wage

$16.00

Required Service
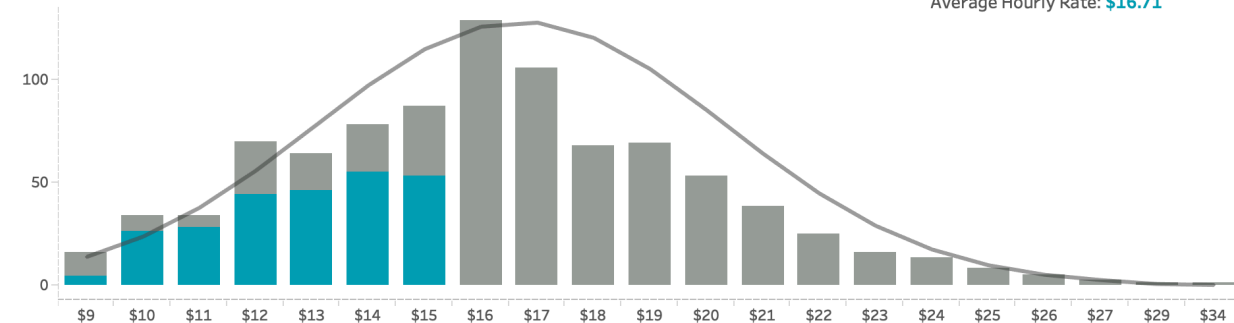
3

Below 256 (27.9%)	**Above** 661 (72.1%)
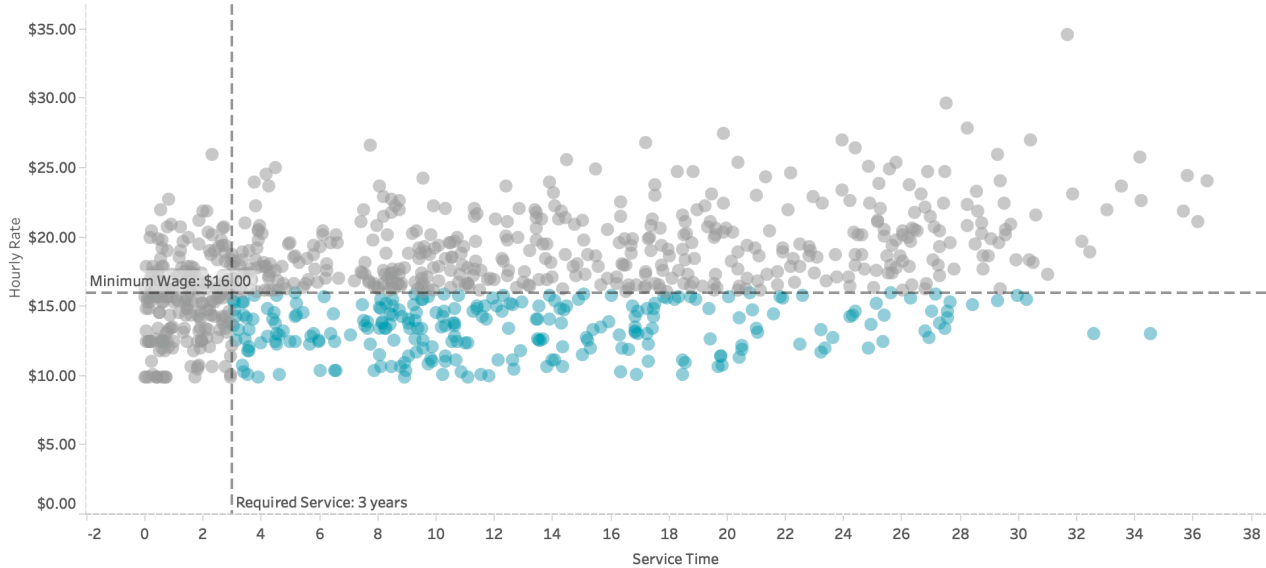
Dollar Impact of Minimum Wage: **$1,792,206**

Employees Below Minimum Wage: **661**

Employees Below Minimum Wage: **256**

Department	Dollar Impact	Employees Below Minimum Wage: 661
Facilities	$42,440	191
Legal	$30,108	6
Logistics	$16,764	38
Engineering	-$38,645	12
Services	-$87,052	309
Information Technology	-$107,696	19
Purchasing	-$116,048	27
Customer Service	-$121,224	28
Operations	-$166,590	35
Marketing	-$189,834	91
Finance	-$198,323	15
Research & Development	-$283,377	39
Human Resources	-$351,142	32
Supply Chain	-$528,309	75

Current Distribution

Average Hourly Rate: **$16.71**

Distribution with Minimum Wage

Average Hourly Rate: **$17.44**