Easy and Stable Deployment of R-based Software with GitHub, devtools, and Docker

Daeyeong Kim, Jinsung Kim, Seonjeong Lee, Hyoseop Lee at Encored Technologies, Inc.

An unexpected upgrade in dependent packages can **break** your production

Our failure: the default operation of `dplyr::distinct` changed

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

expected: `%% distinct(a, b)`
after update: `distinct(data, ..., .dots, .keep_all = FALSE)`

Requirements on **new devtools::install_github**
- Do **NOT** upgrade dependent packages unless it violates version requirements
- Dependent packages are installed as a minimum required version

```
dplyr  
old   new
(House >= old)
```

Reformed **devtools::install_github** to be **conservative** on installing dependencies

Benchmark [pip install](https://pip.pypa.io/en/stable/)

<table>
<thead>
<tr>
<th></th>
<th><code>pip install SomePackage==1.0.4</code></th>
<th><code>pip install SomePackage</code></th>
<th><code>pip install 'SomePackage&gt;=1.0.4'</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>specific version</td>
<td>latest version</td>
<td>minimum version</td>
</tr>
</tbody>
</table>

Reformed `devtools::install_github` works as following:

**At 2016-06-30**

**DESCRIPTION of HanSolo v0.1.0**

- Package: HanSolo
- Version: 0.1.0
- Imports: `dplyr` (>= 0.4.3), `MillenniumFalcon` (>= 0.2.3)

Latest version of dependent packages
- `dplyr`: 0.5.0
- `MillenniumFalcon`: 0.2.3

- **Case 1-1**: initial installation on totally empty environment at 2016-06-30
  - Any packages related to HanSolo are not installed yet
  - Running `install_github("Encored/HanSolo")` would install `dplyr(0.4.3), MillenniumFalcon(0.2.3)` and `HanSolo(0.1.0)`
  - Latest version of `dplyr(0.5.0)` would not be installed since it’s not the minimum required version

- **Case 2**: initial installation on preoccupied environment at 2016-06-30
  - `dplyr(0.5.0)` is currently installed
  - Running `install_github("Encored/HanSolo")` would install `MillenniumFalcon(0.2.3)` only
  - `dplyr` would not be installed since its requirement already being satisfied
  - It would not awkwardly install the minimum required version when requirements are satisfied with margin

**At 2017-06-20**

**DESCRIPTION of HanSolo v0.2.0**

- Package: HanSolo
- Version: 0.2.0
- Imports: `dplyr` (>= 0.4.3), `MillenniumFalcon` (>= 0.3.0)

Latest version of dependent packages
- `dplyr`: 0.7.6
- `MillenniumFalcon`: 0.3.0

- **Case 1-2**: Re-installation on initially deployed environment at 2017-06-20
  - The only unsatisfied requirement is `MillenniumFalcon`
  - Running `install_github("Encored/HanSolo")` would install `MillenniumFalcon(0.3.0)` and `HanSolo(0.2.0)`
  - `dplyr` would not even being considered to install since its requirement already being satisfied

**Iteratively build Docker images using reformed devtools::install_github**

Using reformed `install_github`, any R-base SW can be safely stacked up on a docker image without any side-effects due to unwilling upgrades in dependent packages

ACKNOWLEDGMENTS

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20161210200410).