

# UC San Diego Skillscape Workforce Dashboard Technical Guide

Updated August 26, 2021



An Emsi Creation

UC San Diego  
EXTENSION

CR+E

CENTER FOR RESEARCH  
+ EVALUATION

# Welcome to the UC San Diego Skillscape Workforce Dashboard

## Quick Start Guide

- Begin by reviewing the Quick Start Guide first – this provides the most essential information needed to begin engaging with the tool. It is available in the UC San Diego Skillscape Workforce Dashboard tool and also the accompanying UC San Diego Skillscape Technical Guide.

## Skillscape Technical Guide

- The Skillscape Technical Guide is a PowerPoint deck (currently) that includes the Quick Start Guide and - more importantly - contains expanded documentation to guide and support users. This guide is a work in progress so corrections and comments are welcomed, but it should already be helpful.
- Three sections of the technical guide may be of interest to new users:
  - 1) **Measuring Labor Market Needs** defines the various metrics used in Skillscape to assess market needs
  - 2) **Key Concepts** explains and discusses the central components of Skillscape
  - 3) **Skill Clusters - The rationale** will enrich understanding of the power and potential of skill clusters

## Skillscape Dashboards

- Users should engage Skillscape dashboards in their given order, and in full-screen mode. The dashboards grow in complexity and power from left to right; fundamental concepts and visual cues are introduced first in the simpler dashboards (tabs colored green) before more complex concepts, visuals and capabilities are incrementally introduced in more advanced dashboards (tabs colored blue).
- Red **TIPS** and **NOTES** are included throughout the dashboards to assist new users within the tool, but detailed documentation is now provided in the separate technical guide.

Please direct questions and suggestions for improving Skillscape to Dr. Stuart Muller ([stuart.muller@economicmodeling.com](mailto:stuart.muller@economicmodeling.com))

# UC San Diego Skillscape Workforce Dashboard

Quick Start Guide

UC San Diego  
EXTENSION

CR+E  
CENTER FOR RESEARCH  
+ EVALUATION



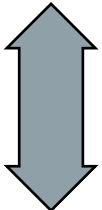
# What is Skillscape?

- Skillscape is shorthand for various Tableau dashboard tools that use Emsi Skill Clusters to explore the “skills landscape” in regional labor markets
- Skillscape is framed to assist higher education in identifying programmatic needs and opportunities
- Skillscape provides skills-based data and analytics for 26 California MSAs, the state of California, and the Nation
- Labor market signals are evaluated through three principal lenses:
  - **Skill Clusters**: statistically derived conglomerations of skills based on the patterns of correlation between skills in job postings
  - **Skills**: includes knowledge, skills, abilities, technologies, etc. from Emsi’s open-source library of 30,000 skills
  - **Career Sub-Areas**: groups of related occupations (SOC codes) that capture broad sectors of the labor market.

# How can Skillscape help?

- Use views with **Skill Clusters** to identify and explore important skills-based “**Cluster Themes**” in the regional labor market demand.
  - Skill Clusters typically provide insights and suggestions at the scale of degree programs or specializations.
- Use views with **Skills** to clarify finer scale needs within and across **Cluster Themes**.
  - **Skills** typically provide insights and suggestions relevant to the scale of specialization, certificates, micro-credentials, badges, continuing ed, courses and lessons.

Scale	Relevant Application of Demand Signals							
Unit of Analysis	Schools & Colleges	Programs & Degrees	Master's	Specialization	Certification	Micro-creds & Badges	Courses	Lessons
Career Sub-Areas	●	●						
Clusters		●	●	●				
Skills				●	●	●	●	●



Coarser

Finer

# Skillscape Orientation

- All dashboards have the same three main filtering options along the top of the view, which directly control the choice of clusters. Changes made in one dashboard apply to all.

**TIP:** Use Step 1 to explore market needs for 24 California MSAs, state of California or the Nation

**TIP:** Use Step 2 to explore market needs for schools & programs

**TIP:** Use Step 3 to explore market needs for BS & MS degrees, specializations & certifications

**STEP 1** Select your region

**STEP 2** Select your target career areas

**STEP 3** Refine your search or view of skill clusters

Region Name

Career Area

Career Sub-Area

Cluster Theme

Filter Clusters by keyword

- Dashboards that also display individual skill information have additional filtering options – shaded blue - on the right-hand side, which enable refining the views of skills. Changes made one dashboard apply to all, unless specified otherwise.

# Skillscape Dashboards

Dashboard Name	Description	Application
Regional Overview	Summarizes supply, demand, salary, and growth projections for the top clusters	What are the top clusters in a region that employers are posting, hiring, and paying for?
Cluster Postings	Summarizes demand in terms of job postings that match to clusters	What are employers posting for and how has the pandemic impacted hiring activity?
Cluster Demand	Summarizes demand in terms of modeled available jobs and talent supply, 5-year new job growth, and salary	What is the actual outlook for jobs, relative to supply, projected growth and possible earnings?
Skill Demand by Cluster	Summarizes demand for skills within skill clusters	What is the demand for specific skills in the region?
Skill Gaps	Offers a view of gaps between skill demand and skill supply	Explore the need for skills within skill clusters in a region
Regional Skill Clusters	Offers 3 different views of clusters that showcase the skills and metadata	Explore regional clusters in greater detail to better understand regional needs
Regional Opportunities	Provides comparative views of clusters with connections to skills and metadata	Identify regional opportunities by comparing clusters

# UC San Diego Skillscape Workforce Dashboard

Measuring Labor Market Needs

UC San Diego  
EXTENSION

CR+E  
CENTER FOR RESEARCH  
+ EVALUATION





# Cluster Postings

- **Job postings** provide a simple yet effective measure of the market's need. Emsi matches job postings to clusters by comparing the alignment between skills requested in the posting and skills in the cluster.
- **2019 Matched Postings:** Number of active postings matching the cluster in 2019
- **2020 Matched Postings:** Number of active postings matching the cluster in 2020
- **COVID-19 Impact:** Percent change in matched postings activity due to the pandemic
- Since many postings are for menial jobs or simply very brief and mention few skills, not all are matched to clusters. Larger MSA's may match over three-quarters of postings to clusters but smaller MSAs may match fewer than half.
- Even so, there are always many more job postings than there are actual openings to fill. Emsi de-duplicates job postings but this is challenging for remote jobs, and many postings simply go unfilled.
- Job postings provide a window into interests and intensity of hiring activity but are not a reliable quantifier of actual demand. The Regional Cluster Demand view introduces estimates of demand that use Bureau of Labor Statistics measures of actual job openings.

# Cluster Metadata Postings

- Though not prescribed up front, each cluster is ultimately the result of a unique sample of job postings.
- Each job posting is tagged with a company name, education levels, industry code (NAICS), occupation code (SOCS), and job title, and the top aggregate results of each are summarized in the Cluster Metadata views.
- Cluster metadata counts are useful signals of relative importance and play an important role in the interpretation of clusters, providing valuable context about the most prominent drivers of the cluster.

# Cluster Demand & Supply

- Cluster **Supply** is a modeled estimate of available talent that matches the cluster
  - Emsi models estimates of Supply by synthesizing professional profiles data with Bureau of Labor Statistics (BLS) data from the Current Employment Statistics (CES) and the Quarterly Census of Employment and Wages (QCEW) surveys. Profiles are matched to clusters and provide the shapes of the talent pool in terms of skills. BLS data help calibrate those shapes to the scale of the actual workforce.
- Cluster **Demand** is a modeled estimate of annual job openings that match the cluster
  - Emsi models estimates of Demand by synthesizing scraped job postings data and Bureau of Labor Statistics (BLS) data from the Job Openings & Labor Turnover Survey (JOLTS). Postings data provide the shapes of demand in terms of skills. BLS data help calibrate those shapes to the scale of JOLTS openings.
- **Demand/Supply Ratio** indicates a regional talent gap for the cluster when  $> 1$  and a talent surplus when  $< 1$ .

# Cluster Job Growth & Salaries

- **Job Growth (2019-2024)** is an estimate of the number of new jobs in this cluster over 5 years.
- Job growth for a cluster is estimated using Emsi's occupational growth projections and metadata about the distribution of occupations in the job postings that created each cluster.
- **Salary** is an estimate of median annual earnings
- Salary is estimated from salary data sourced directly from job postings and from BLS Occupational Employment and Wage Statistics (OEWS).

# Skill Demand & Supply

- Skillscape provides estimates of Demand for skills within each cluster, which can be aggregated across multiple clusters to quantify demand for specific skills across broader sectors of the market or larger regions.
- Individual **Skill Demand** is estimated from the matched postings and modeled cluster demand. The distribution of skills in matched postings determines the shape of demand, which is then scaled according to the modeled magnitude of demand for each cluster.
- Individual **Skill Supply** is estimated in the same fashion but used matched profiles to determine the shape of their distribution.

# UC San Diego Skillscape Workforce Dashboard

Key concepts

UC San Diego  
EXTENSION

CR+E  
CENTER FOR RESEARCH  
+ EVALUATION



# Emsi Skill Clusters

- Emsi analyzes the skills data in job postings to surface the most statistically meaningful groupings of skills.
- Clusters can be seen as a new kind of framework that describes “market roles” based on the skills being requested. This skills-based framework serves an analogous function to the Standard Occupational Classification System (SOCS) and O\*NET system in that it is a helpful way to describe, classify and quantify kinds of market demand.
- However, instead of an inherited system of pre-defined occupations, only one of which a job can be tagged to despite the widespread overlap in skills between many occupations in practice, Emsi’s “skills-up” approach allows the skills being asked for in job postings to define the market roles – i.e. skill clusters.
- Clusters represent the most statistically important kinds of need in sectors of the labor market, as determined by the skills mentioned in postings and by their patterns of association with each other.
- Clusters do not necessarily describe the entire labor market in a region. Rather, they are the most important and distinctive aspects of the labor market demand and help us direct attention and resources accordingly.

# Skill Cluster Regions

- Emsi performs skill cluster analyses at three scales – Metropolitan Statistical Areas (MSAs), States and Nationwide
- Each scale of analysis surfaces the distinctive clusters that emerge at that scale, so it is often revealing to compare the clusters that arise in the same sector but different regions.
- Data availability is an important limiting factor in cluster analyses so smaller regions – particularly small MSAs – may surface fewer clusters, which may lead to fewer matched job postings.
- Large MSAs can match over three quarters of postings to a cluster in their region, but smaller MSAs may match less than half. Low numbers of matched job postings do not necessarily lead to lower modeled estimates of Cluster Demand, but the shape of skill distribution (which relies on matched postings – see Cluster Demand & Supply slide) may inherit some biases from this.



# Cluster Themes

- The output of the cluster analysis is groupings of skills, which are not very practical to work with until digested and named.
- Emsi has used a combination of intensive human curation and savvy statistics to provide most clusters with a name that describes the cluster – this is the Cluster Theme.
- Cluster Themes have been *suggested* based on a general review of the skills in the cluster, taking into account the most prevalent job titles and occupations that contributed to the cluster.
- Since clusters are “aggregate” representations of groups of skills needed in similar or related roles, a subject matter expert may be able to discern multiple specific and overlapping themes within a cluster. In some cases, these may not fit well with the intentionally more general Cluster Theme that Emsi has suggested.
- Cluster Themes should therefore be seen as descriptors of general kinds of “roles” that the market is seeking, and not treated as formal definitions.
- Themes are also an indispensable tool when exploring market demand at the cluster scale; these filters are shaded red in the dashboards.

# Career Sub-Areas & Career Areas

- A key advantage of Emsi's skill clustering is the ability to find patterns in skill demand and correlation across diverse but related jobs. Emsi has created 21 Career Areas and 65 Career Sub-Areas by grouping related occupations (SOC) together.
- Tens of thousands of job postings are randomly sampled from each Career Sub-Area and analyzed to surface the most statistically meaningful groupings of skills – the skill clusters – in each Sub-Area.
- Career Areas and Sub-Areas are crucial filters for navigating and working with clusters.
- A table showing the full mapping of all SOC occupations to Career Areas and Sub-Areas is provided.
- The Career Area and Sub-Area filters are shaded orange in the dashboards.

# Emsi Skills & Skill types

- The Emsi Skill Library comprises approximately 30,000 **Skills** drawn primarily from natural language processing of job postings.
- Emsi uses “skills” in a very general way that encompasses knowledge, skills, abilities, technologies, etc.
- With over 30,000 skills it becomes necessary to group them in meaningful ways that are relevant across clusters.
- Emsi has defined 15 “**Skill Types**” that categorize skills more generally, e.g., Marketing & Communications.
  - Skills can only be tagged to one Skill Type.
- Skill Type filters are shaded blue in the dashboards

# Skill Clustering Methods

- Emsi aggregates Standard Occupation Codes (SOCs) into Career Areas and Career Sub-Areas (see Reference Tables dashboard for SOC to Emsi Careers mapping). Skills data in job postings are analyzed at the scale of Career Sub-Areas for specific regions, using statistical methods that account for the relationships between skills in job postings. The result is sets of skill clusters for a given region and Career Sub-Area, which reflect the kinds of roles that employers are posting for based on the skills they seek. Postings are drawn from the most recent two years, and clusters are updated approximately once every 6 months.
- Each skill in a cluster has an associated Skill Score (the Standard Score in statistical jargon) for that skill in that cluster. This score indicates the degree of significance of a skill to a given cluster. A higher score suggests the skill is of "core" relevance to the cluster, whereas a lower score suggests the skill may be more "niche."
- Some common skills (e.g. Communications) are excluded from the cluster analysis because they are so ubiquitous that they tend to dilute or confound the important relationships between other skills, while not adding very much of value in and of themselves.

# Emsi's Clustering Approach to Skills Data

- Analyze skills based on their relationships with each other
- Method: Corex - factor analysis of skill correlations
- X-axis: Skill Demand
- Y-axis: Skill Score
- Skills cluster into "roles" based on market demand
- Skills contextualized by clusters & adjacent skills

## THE NATIONAL DATA SCIENCE SKILLS CLUSTER

Skills that characterize data science



SKILL HAS HIGH CORRELATION TO FIELD

SKILL APPEARS IN HIGH NUMBER OF JOB POSTINGS

**Horizontal axis:** Shows the frequency that each skill occurs in the job postings.  
**Vertical axis:** Shows the correlation between the skill and the cluster it belongs to. The higher the correlation, the more important the skill is in explaining the cluster.

# Emsi Skills Library

- How does Emsi update skills? The Emsi skills taxonomy is a comprehensive database that encompasses nearly thirty thousand unique job skills and certifications. Every two weeks, Emsi releases a new version of the taxonomy to add new and emerging skills, modify existing skills, and remove obsolete skills. We rely on feedback, research, and machine learning to evaluate and refine this database.
- Adding skills: Emsi adds skills to the database frequently based on feedback and research. For potentially ambiguous words and phrases, we use machine learning classifiers to recognize new skills in job postings and profiles. Because this process involves careful curation and training, it can take several weeks or months to add a new skill. However, many skills are unambiguously identifiable and do not require machine learning; with a bit of research, these can debut in the next release.
- Modifying skills: Because of the nature of the taxonomy and the evolving language of job postings and skills, some skills require modification to improve skill recognition. After detecting an issue with a skill, we evaluate it to understand the problem. This informs the next steps of modification. Skill improvement can involve quick updates (e.g., name changes, consolidations, or whitelisting) or further training of the underlying machine learning model. Sometimes skills cannot be updated because there is not a good way to do so without losing valuable (albeit noisy) information.
- Removing skills: Current work to curate the database involves removing outdated or extraneous skills. Before removing skills from the taxonomy, we assess them to understand why they are being tagged. Sometimes, a skill name-change or quick update can solve the problem. Other times, skills are irrelevant or obsolete and we can remove them from the database in the next release.
- Data Integrity: The overall integrity of the skills taxonomy is maintained via the use of unique skill identifiers. Users of our skill taxonomy are not required to keep pace with our regular release cycle as we do not reassign or reuse identifiers. Due to this, consumers of our APIs can safely pin to older versions of the taxonomy and simply ignore skills which they do not recognize.

# UC San Diego Skillscape Workforce Dashboard

The Rationale Behind Skill Clusters

UC San Diego  
EXTENSION

CR+E  
CENTER FOR RESEARCH  
+ EVALUATION





Each job posting is tagged with skills...

...but what to do with 30,000 different skills scattered across millions of job postings?



# Standard Approach to Skills Data

- Slice the job postings up according to tagged occupations or job titles, and then sum the frequencies of individual skills

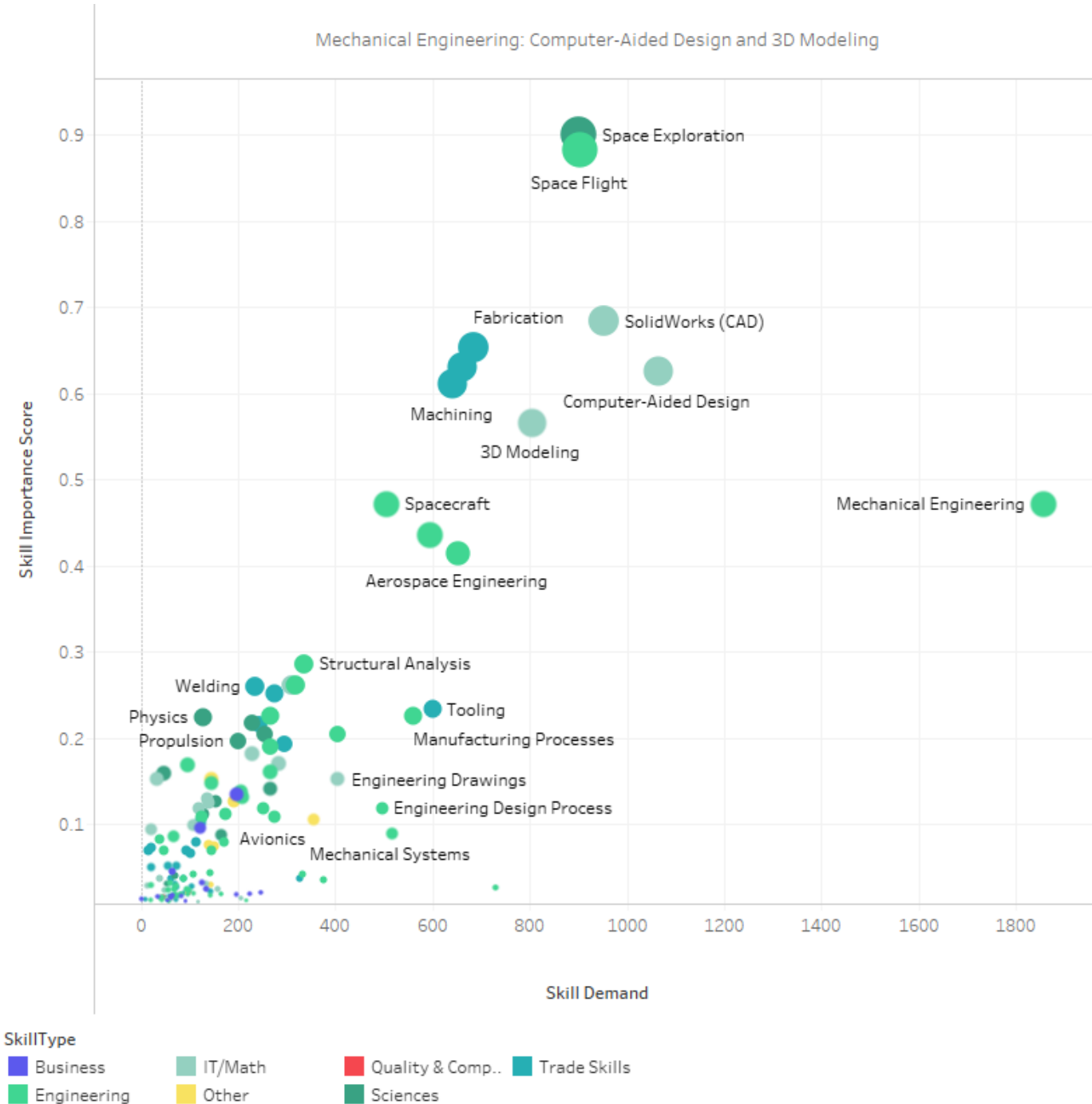


# But...

- This works well enough if you have an occupation code specific enough to be relevant – many are too narrow or too broad - or a very distinctive job title in mind – though Emsi has 75,000 job titles and most jobs could be described by a dozen different titles.
- If you search more generally, the skill results tend to be more general, with the most commonly referenced skills – often the least-interesting – obscuring the significance of less obvious skills – though often more important.
- There is also little way to know how skills relate to each other when all you have is a list counting the number of times they individually showed up. Showed up with what else? The possible combinations are too complex to manually analyze.
- And yet, these combinations of skills are the crux of the matter.
- Traditional, well-understood job roles have become fuzzier; cross-disciplinary work has become ubiquitous (especially with regards to technology in our jobs); and employers are opening-up to new modes of education and experience – demand is more mixed, and supply is more mixable. The ingredients are skills.
- We create large samples of postings from groups of related occupations – Career Areas – and extract all the skills from all the postings. We then analyze those data to see how skills co-occurred with each other and with groups of each other. The end-product is a cluster of skills that represent a statistically significant pattern of demand.
- The y-axis is a measure of each skill’s “weight” in the cluster – from the more core skills (approaching 1) that tend to be pervasive to the more peripheral skills (approaching 0) that may be cross-disciplinary or specializations.
- The x-axis is an estimate of the number of postings requesting that skill for jobs of that cluster’s sort.
- Clusters represent an amalgam of jobs, so cluster names should be seen as interpretations of clusters, not definitions. Different perspectives may interpret a cluster differently based on different priorities, but we provide metadata about the postings that generated each cluster to support that interpretation – the top job titles, occupations, companies, education requested and industries. Refining cluster names can be very helpful.

# So...

- Emsi built a process that puts skills first and builds meaning from there.
- We use big data statistics to analyze the distributions of skills and of their relationships with each other and find statistically significant patterns in the demand **at the skill-scale** – the “skill clusters.”
- The y-axis is **Skill Importance**: a sign of each skill’s “weight” in the cluster. Skills close to zero may not be common but may capture important intersections and specializations.
- The x-axis is **Skill Demand**: an estimate of the annual number of jobs requesting that skill for roles in that cluster.
- Skills are also categorized at 2 scales:
  - **Skill**: Individual skill/ability/knowledge area
  - **Skill Type**: A broad kind (e.g., Business)



Occupation codes are still useful and important, but classification systems always have a fixed scale that limit higher resolution understanding of what the market truly wants – skills.

Skill clusters are a bottom-up – “skills-up” – approach that allows us to define roles and explore relationships between jobs at the scale of skills.

### Taxonomic Approach

Related by category

#### 19-0000 Life, Physical, and Social Science Occupations

##### 19-1000 Life Scientists

19-1011 Animal Scientists

19-1012 Food Scientists and Technologists

19-1013 Soil and Plant Scientists

##### 19-1021 Biochemists and Biophysicists

19-1022 Microbiologists

19-1023 Zoologists and Wildlife Biologists

##### 19-1029 Biological Scientists, All Other

##### 19-2000 Physical Scientists

19-2011 Astronomers

19-2012 Physicists

19-2021 Atmospheric and Space Scientists

### Skill Shape Approach

Categorized by relationship



# Skills-based Multi-scale Labor Market Demand Signals

